JULY 1961
Volume XVII Number 3

Journal of Clinical Psychology

IN THIS ISSUE:

PSYCHOLOGICAL CLINIC FROM NEW JERSEY

EDITORIAL BOARD

FREDERICK C. THORNE, *Editor*
Editorial Office: 5 Pearl Street
Brandon, Vermont

JERRY W. CARTER, JR.
U. S. Public Health Service

HOWARD F. HUNT
University of Chicago

WILLIAM A. HUNT
Northwestern University

ELAINE F. KINDER
46 Kinney St., Piermont, N. Y.

MAURICE LORR
Veterans Administration Washington, D. C.

BERNARD SAPER
N. Y. State Department of Mental Hygiene

SAUL B. SELLS
Texas Christian University

ROBERT I. WATSON
Northwestern University

PETER THORNE, *Business Manager*

Journal of Clinical Psychology
5 Pearl Street - - Brandon, Vermont

Ready in September . . .

PERSONALITY

AN ECLECTIC CLINICAL APPROACH

Frederick C. Thorne, M.D., Ph.D., Editor

Journal of Clinical Psychology

- This book presents a comprehensive clinical approach to personality status based upon an eclectic study of the factors actually organizing integration.
- The model of personality integration postulated in this book is derived from the actual raw data of personality status rather then upon hypothetical considerations. The approach is derived from actual clinical study of the person/organism.
- A series of basic postulates are presented concerning factors actually integrating behavior at any moment in personality time and space.
- An operational analysis is presented of the contributions of the various schools of psychology and psychiatry within the general eclectic orientation.
- A broad spectrum approach which considers all levels of hierarchical personality organization including the global pattern of the person working out his life. Suitable for all classes in the psychology of personality.

REGULAR PRICE: $6.00

SPECIAL PREPUBLICATION OFFER BEFORE OCTOBER 1, 1961: $5.00

Companion Volumes Outlining a Comprehensive Eclectic System of Clinical Practice

DIAGNOSIS ● Principles of Psychological Examining
Frederick C. Thorne, M.D., Ph.D. PRICE: $7.50

THERAPY ● Principles of Personality Counseling
Frederick C. Thorne, M.D., Ph.D. PRICE: $7.50

PRACTICE ● Clinical Judgment
Frederick C. Thorne, M.D., Ph.D. PRICE: $7.50

ORDER NOW

Journal of Clinical Psychology
5 Pearl Street
Brandon, Vermont

JOURNAL OF CLINICAL PSYCHOLOGY

VOL. XVII JULY 1961 No. 3

CONTENTS

PSYCHOLOGICAL CLINIC FROM NEW JERSEY

Co-editor's Introduction. HENRY P. DAVID - - - - - - - - - - 219

Outpatient Treatment of a Seriously Disturbed Child in a Child Guidance Center. FRANCES SEIDMAN - - - - - - - - - - - - - - 220

Analysis of Interpersonal Relationship as an Aid to Understanding Family Dynamics: A Case Report. BERNARD GUERNEY, JR. and LOUISE GUERNEY - 225

Individual Therapy with a "Defective Delinquent". ERWIN FRIEDMAN - 229

Problems of Psychological Differential Diagnosis Associated with Hypothyroidism: A Case Study. REBECCA SNYDER - - - - - - - - - 232

Comparison of Leiter, WAIS and Stanford-Binet IQ's in Retardates. JOHANNES C. BRENGELMANN and JOSEPH T. KENNY - - - - - - - - 235

Anxiety as "Effect of Uncertainty": An Experiment Illuminating the OA Subtest of the WAIS. EILEEN S. LANFELD AND D. B. SAUNDERS - - 238

An Evaluation of Rorschach Reliability Through the Alternate-Response Method. HARRY STEIN - - - - - - - - - - - - - - - - - 241

Homosexuality and Some MMPI Measures of Masculinity-Femininity. BERNARD S. AARONSON AND HOWARD R. GRUMPELT - - - - - - 245

Eysenck's "Stratification" Theory and the Questionnaire Method of Measuring Personality. C. M. FRANKS, E. A. HOLDEN and M. PHILLIPS - - 248

CLINICAL PSYCHOLOGY IN MENTAL HOSPITALS

A Sample Survey of Mental Hospital Patients. E. I. BURDOCK, LEAH GLASS, ANNE S. HARDESTY AND YVONNE BECK - - - - - - - - - - 253

A Behavioral Disturbance Index for Psychiatric Patient and Ward Disturbance. JAMES L. FRAMO AND ARTHUR M. ADLERSTEIN - - - - - - 260

Consultation vs. Therapy in the Psychological Treatment of NP Hospital Patients. MILTON B. JENSEN - - - - - - - - - - - - - - 265

STUDIES OF SCHIZOPHRENIC PROCESSES

Empirical Syndrome Groupings of Schizophrenic Hospital Admissions. WILSON H. GUERTIN - - - - - - - - - - - - - - - - - 268

Validities of Three Visual Tests for Differentiating Organics from Schizophrenics and Normals. BERT R. SAPPENFIELD and ROBERT J. RIPKE - - 276

The Use of Drawings of Hands as a Projective Medium for Differentiating Normals and Schizophrenics. EDWIN E. WAGNER - - - - - - - 279

VALIDATION STUDIES: OBJECTIVE TESTS

Short Forms of the WISC for Use with Emotionally Disturbed Children. RICHARD ENBURG, VINTON N. ROWLEY and BETH STONE - - - - - 280

The Ceiling of the New Stanford-Binet. WALLACE A. KENNEDY, HAROLD MOON, WILLARD NELSON, RONALD LINDNER and JACK TURNER - - 284

Handedness and Digit Symbol Performance. RICHARD J. BONIER AND CHARLES HANLEY - - - - - - - - - - - - - - - - - - 286

A Validity Study of Tamkin's "MMPI Scale Measuring Severity of Psychopathology. DANIEL R. LUTZKER - - - - - - - - - - - - - 289

CONTENTS—Continued

An Abbreviated Qualitative Score for the Porteus Mazes. JOHN R. BARRY, LOUISE EVERSTINE and JOHN P. KLEMAN - - - - - - - - - - - 291
The Applicability of an American Delinquency Scale to Japanese Subjects. T. SHIRASA and T. AZUMA - - - - - - - - - - - - - - - 291

NORMATIVE STUDIES: OBJECTIVE AND PROJECTIVE TESTS

Visuoconstructive and Verbal Responses in Chronic Brain-Damaged Patients and Familial Retardates. EARL E. BALTHAZAR, RONALD E. TODD and DON H. MORRISON - - - - - - - - - - - - - - - 293
Standardization of the Graphoscopic Scale: The Content of Children's Drawings. JUSTIN PIKUNAS and HUGH CARBERRY - - - - - - - 297
Structured Personality Testing in the Aged: An MMPI Study of the Geriontic Population. WENDELL M. SWENSON - - - - - - - - - - 302
Rorschach Responses of Children Exhibiting Psychogenic Auditory Symptoms. FRANK KODMAN, JR. and JERRY W. WATERS - - - - - - - 305
Nudity in Human Figure Drawings. GARY M. FISHER - - - - - - 307

Performance Changes in Repeated Judgments of Schizophrenic Verbal Responses. STANLEY BLUMBERG - - - - - - - - - - - - - - 308
Below Chance Performance on a Verbal Test Using Paired Concepts. E. FRANCESCO - 312
Social Desirability and the Type of Hostility Expressed on the TAT. MARVIN REZNIKOFF AND ADELAIDE DOLLIN - - - - - - - - - - 315
A Longitudinal Comparison of Intelligence Involving The Wechsler-Bellevue I and WAIS. DONN R. DUNCAN AND ALBERT M. BARRETT - - - 318
Psychological Effects of Isocarboxazid and Nialamide on a Group of Depressed Patients. VASSO VASSILIOU AND H. E. HIMWICH - - - - - 319
A Cross-Validation of Leary's "Public" Communication Level. WALTER G. KLOPFER - 321

SELECTED CASE STUDY

Changes in Essential Hypertension over Time: A Factor Analytic Study G. A. MILTON - - - - - - - - - - - - - - - - - - - 322

Editorial Opinion - 326

GENERAL INFORMATION

The *Journal of Clinical Psychology* is an independent journal dedicated to the advancement of the clinical method in psychology. Although primarily a scientifically oriented professional journal limited to publication of original research reports and authoritative theoretical articles, it aims to foster the promotion and expansion of clinical psychology as an applied science. Special sections in each issue will be devoted to news, comments, and letters to the editors on matters pertinent to the profession. Communications and articles submitted for publication are welcome and should be addressed to the Editorial Office, 5 Pearl Street, Brandon, Vermont.

MANUSCRIPTS must be typewritten, double-spaced with good margins, on one side of the paper only. The publishers will not be responsible for manuscripts lost in the mails or through unavoidable accident. Retain a carbon copy after sending the original to the editor. It is requested that all articles be as brief as consistent with clarity. All manuscripts are subject to editorial modification.

REPRINTS are furnished on order only and must be requested on the order blank provided when galley proofs are returned. Change of address notices should be sent one month before moving and should include old and new addresses. Undelivered copies will be remailed only at subscribers' expense.

The *Journal of Clinical Psychology* is published quarterly, appearing in January, April, July and October, at Rutland, Vermont, by Frederick C. Thorne, M.D. Subscription price $10.00; Canadian subscriptions $10.50; foreign subscriptions $11.00 including postage; single copies $2.50. Second-class postage paid at Brandon, Vermont and at additional mailing offices. *Title registered, U. S. Patent Office.* Editorial and Business Offices, 5 Pearl Street, Brandon, Vermont.

Copyright, 1961, Frederick C. Thorne, M. D.

SPECIAL FEATURE

PSYCHOLOGICAL CLINIC FROM NEW JERSEY

It gives me much pleasure to introduce a series of papers representative of current clinical research in Community, State, and Federal institutions and educational centers in New Jersey. An idea suggested by my colleague, Bernard S. Aaronson, sparked this collection. Clinical psychological activities in the New Jersey State Department of Institutions formally began in 1919 when J. E. Wallace Wallin left the Vineland Training School to establish a psychological laboratory at the New Jersey Village for Epileptics, the site now occupied by the New Jersey Neuro-Psychiatric Institute. Much has happened in the dynamic fifty years that followed, as recorded in the special 1960 commemorative issue of the New Jersey *Welfare Reporter*, a copy of which is available on request to my office.

The first four papers in this series relate clinical case material, reflecting the broad scope of New Jersey activities and close ties among State and partially State supported agencies and Federal institutions. Frances Seidman reports on long-term out-patient therapy with an autistic youngster in a community child guidance center. The importance of family dynamics and the need to develop appropriate psychological techniques of assessment are clearly apparent from the Guerneys' study of a disturbed young girl and her parents in the Rutgers University Psychological Clinic. Next, Erwin Friedman shows what can be done in individual therapy with a "defective delinquent", rejected by his family and relegated by society to a State colony. Rebecca Snyder completes the cycle with a diagnostic study of an adult veteran, initially diagnosed paranoid schizophrenic, whose psychosis was eventually established to be on a metabolic basis.

The next group of five papers focuses on techniques and problems of evaluation. J. C. Brengelmann and Joseph T. Kenny note the wide intelligence quotient variations in retardates as reported by different tests, and provide experimental evidence indicating that the higher intelligence the higher Performance IQ, and the lower intelligence the higher Verbal IQ. Eileen Lanfeld and David Saunders consider anxiety in terms of an "uncertainty effect", and show how a sentence arrangement test can be developed to measure what is currently obtained from the WAIS Object Arrangement Subtest. Harry Stein explores Rorschach reliability and offers several suggestions for its measurement. The MMPI is used by Bernard S. Aaronson and Howard Grumpelt to substantiate the clinical observation that some male homosexuals have a primarily masculine orientation and that masculinity-femininity is not a unitary dimension. Finally, C. M. Franks, E. A. Holden, and M. Phillips demonstrate the possible dangers of using self-rating personality questionnaires as measures of objective behavior without more extensive consideration of the population under observation.

HENRY P. DAVID, GUEST EDITOR
New Jersey State Department of Institutions and Agencies

OUTPATIENT TREATMENT OF A SERIOUSLY DISTURBED CHILD IN A CHILD GUIDANCE CENTER

FRANCES SEIDMAN

*Child Guidance Center of Mercer County**

Diagnostic Evaluation

Leslie was four when she first came to the Center. She was large for her age, sad-eyed and dull, and she left her mother with neither a backward nor a forward glance. Leslie had been referred by the pediatrician because of hyperactivity and possible mental retardation, and she gave the immediate impression of acute emotional disturbance. Her general functioning was typified by an unrelatedness and a fear of being close to anyone. In the playroom she was not hyperactive; but she was not able to concentrate on any one activity. Her speech was infantile and suggestive of the schizophrenic process. She used third person pronouns in referring to herself and others in the room, and used echolalia and clang associations. In some respects, Leslie appeared to meet Bender's description of the schizophrenic child who is pseudo-defective and autistic. There were more hopeful signs, however, in her normal motor development, her low average social maturity as measured by the Vineland Social Maturity Scale, and in her ability to make some object relations. For example, while she did not directly recognize the psychiatrist's presence she placed her coat on his and said, "on his coat". In the psychological evaluation Leslie was able to do some of the picture vocabulary and several non-verbal tasks at about the three-year level. She scribbled when asked to draw a man and then drew a large black spider. She kept asking the psychologist to draw spiders, wasps, and webs for her.

Leslie's mother told of a wanted child and an easy pregnancy. She said, however, that shortly after the birth she reacted with intense fear to the doctor's comment, "I think she's okay". She could hear only the word "think" and she was sure that something was wrong. She recalled her own tension and inability to hold the child. Leslie "felt like a board" when she tried to feed her. It was with relief that she allowed her sister, who was a registered nurse, and other members of her large family to take over much of the responsibility for Leslie. When Leslie was five-weeks old she was hospitalized for a six-day period for projectile vomiting and diarrhea. She went home with her aunt rather than her mother. She was both weaned and toilet trained by relatives on some of her frequent visits away from home. She was handed like a doll from person to person, "spoiled" and "loved" by everyone. Leslie had three adult half-siblings, children of her father, living at home. She was never fed by the same person twice in a row and "never knew what bed she was going to wake up in". At about a year, Leslie was reported to have spoken single words and to be a happy baby. Then at fifteen months, she stopped talking. She was thought to be deaf since no one could "get to her". The withdrawal was accompanied by hyperactivity and restlessness. Medical check-ups, including the EEG, revealed no physical problem. When Leslie was three, her brother Roy was born. At four she had a mild case of mumps, her only physical illness.

An attempt to evolve a suitable treatment plan for a child like Leslie raises questions that cannot be answered with certainty. How severe is the disturbance? What is the treatment of choice? How long will it take? What does the future hold? Despite current theories about the preference for residential treatment for the severely disturbed child, the diagnostic team at the Center proposed a plan for out-

*The Child Guidance Center of Mercer County at Trenton, N. J., is an A. A. P. C. C. approved Clinic that uses the characteristic child guidance team approach. Following a diagnostic period and family conference, the parents are usually seen for weekly therapeutic interviews by a psychiatric social worker. The child has weekly therapy sessions with a child psychiatrist or a child psychologist. Appreciation is expressed to Charles R. Swift, M. D., Director of the Center, who supervised the therapy and took ultimate responsibility for the total treatment program.

patient therapy. This was done with full knowledge that such a plan would require more treatment hours over a longer period of time than is accorded the average treatment case. It was speculated that Leslie would need help, perhaps indefinitely and that it would be well worth the expenditure of Center time in order to study the value of such a continuous program. Prognostically, Leslie was seen as having fair chances for making gains because of the cues for relating that were noted in the evaluation. A summary of the on-going treatment program is presented at this point in order to try to understand how and why it appears to be helping Leslie. The focus of this paper is on the child, who has had 131 hourly treatment contacts over a four-year period. During the first year she was seen regularly, once and then twice a week. Later, sessions were spaced and for the past year she has been seen about once a month. The concurrent social work sessions for the mother are not being studied. The father was interviewed twice in the intake process.[1]

Treatment

Treatment for Leslie began when she first came to the Center. When the therapist met Leslie and her mother in the waiting room, mother reported that Leslie "could hardly wait" to play with the doctor.[2]

T-1. She says my hand is very little and puts her hand in mine . . . she repeats in a far-away voice "little hand Dr. Seidman, little hand".

Anxiety over body was to be a constant theme in treatment. One of the goals that evolved early was an attempt to help Leslie to develop a sense of self—to know her own body boundaries without needing to withdraw completely. Here, as in other areas of work with her, the way was shown by the child. She gave the cues, and the therapist tried to meet her needs. For example, Leslie frequently asked the therapist to draw for her, as she did at the time of evaluation. The therapist offered to draw her, instead of a spider and drawing Leslie became a regular part of treatment.

T-3. She asks me to paint Leslie, and I do so, naming the body parts as I paint. She smiles happily and wants me to paint the sore on her arm. She asks, "Who is Leslie?" and I point to her. She echoes this and then points to me and identifies me . . . She goes back to painting and herself makes a "little Leslie". This is a copy of mine and she seems confident as she makes and names her hair, neck, etc.

T-8. Then she asks me to paint one of Leslie and I do, reciting the little story that goes with it. Then comes the real surprise. She says she will make one of Dr. Seidman and she begins to make a painting of me.

T-12. Today for almost the first time she looks directly in my face and asks what color my eyes are.

T-13. She says, "I want to wear the blue dress in this school that looks cute." She looks closely at me and questions the color of my hair, as well as noting I'm wearing earrings. She says, "What color is your hair Dr. Seidman?"

Almost as if now that she knew who she was and who the therapist was, she began to show clearer signs of identifying with the therapist.

T-26. She comments on her own clothes. She doesn't want to wear buttons because Dr. Seidman isn't wearing any buttons. She wants lipstick like Dr. Seidman.

Sometimes the cues were not as clear, and Leslie's needs could only be inferred. When Leslie seemed content to be alone almost for hours on end, the therapist insisted on doing things for her and doing things with her. Sometimes, however, when the therapist tried to shift the emphasis from some compulsive activity to something that seemed more "healthy", Leslie refused to participate, again showing the way, or at least setting the pace.

T-23. I try to lure Leslie into a ball rolling game; but this does not have meaning for her, and she insists on going back to the clay.

---

[1]Leslie's father was killed in an automobile accident about five months after treatment began.
[2]Movement in therapy will be described from excerpted notes taken from the case record.

Symbolic regression was encouraged by the therapist in response to Leslie's revealing of real basic need for nurturance in the first session.

T-1. I introduce the rubber cow family, and Leslie identifies the mommy and baby. She is concerned over the milk—"mommy has milk. Daddy milk" . . . Leslie drinks from the cow's teat. I recognize the hunger and show her the baby bottle that has "milk" in it too. She drinks a mouthful but goes back to preoccupation with the clay mouse. She sucks the tail and says it tastes good.

During the second session Leslie was coaxed onto the therapist's lap with an offer of hard candy to suck. Gains in her ability to use body contact were made very slowly. Leslie fought contact; then she seemed to enjoy it. Being a baby and "playing baby" was to be a steady theme for a long time. At first she identified herself with her younger brother and later was able to enjoy the regressive play as herself.

T-4. She comes to me for candy and climbs on my lap. I feed her candy, and she slips off quickly. Then she accidently drops a piece, slides off, gets it and crawls back . . . . There is no more candy, and I offer her a drink out of the bottle. She snuggles on my lap and drinks it. "Be your little Roy."

T-6. Leslie asks for candy and sits on my lap in a resisting way - she scrambles off as quickly as she can.

Leslie's decision to enjoy the contact came slowly.

T-30. With only a little bit of help, she becomes a baby. She gets onto my lap and demands to be fed with bottle and "baby candy" . . . Leslie is bathed, diapered, burped, and fed over and over again . . . This is the session I have been waiting for. This type of regression is inevitable and exactly what Leslie needs. If she can get some nurturance out of this experience, it is almost possible that she might be willing to give up autism as a defense.

T-37. She spends a good part of the time sitting on my lap, sucking on the bottle and just being cuddled . . . There is much less preoccupation with the symbols of being a baby, toileting, etc. . . .

A plan to involve mother more closely in Leslie's therapy grew out of mother's expressed concern over Leslie's playing baby, and the hope that more warmth and relatedness could be kindled between mother and child. It was to be almost a re-teaching or a re-living process for them both. Therapy was planned for twice-a-week with the mother's staying in the playroom on alternate sessions.

T-31. Her mother seems capable only of staying remote and then of asking Leslie to identify the animals. She responds quickly to my suggestion that she play real mommy with Leslie and take her on her lap and give her the bottle.

T-35. Leslie wants to be fed almost immediately and I suggest she go to her mother and her mother will give her candy.

During sessions with the mother, Leslie again began to react with withdrawal and to save meaningful material for the alternate sessions.

T-48. The sessions are more like nursery school.

Once a week sessions were resumed and Leslie was more free in the playroom. Her experiences with her mother were useful, however, since mother and child now sat close to each other in the waiting room and mother reported that Leslie was being more affectionate at home and that she herself could give love more openly to Leslie and Roy.

Growing up continued in the treatment sessions as well as at home.

T-102. As Leslie chews a piece of candy, she says, "The baby is cutting teeth". Giving up the comfort of being a baby is not easy. Leslie says, "Tell the little baby a story." and, "Sit on the baby's bed." At the end of the session, Leslie commands, "Teach me to walk Dr. Seidman . . . ."

T-131. Leslie sits on my lap for just a moment "Remember when I played baby?"

Leslie suffered severe trauma early in the therapy period. Leslie was able to use the help of the therapist. Following the accidental death of Leslie's father, Leslie

was brought to the Center by her aunt. Mother was described as being in constant mourning and unable to answer Leslie when she called.

T-18. Leslie keeps repeating in an anxious tone "Mommy wants to bring you" and I said that mommy did want to, but that she was sad because of daddy. Leslie asks, "Where's Leslie's daddy?" Leslie wants to make a painting and begins to paint black splotches on the paper. "Pretty black." As she paints she says, "Mommy wants to take me this time. She's sad and sick. She wants to come, she's sad." Just before the time to leave, Leslie says, "Let's take my hand and go downstairs."

T-19. Leslie is extremely regressed—she is pre-occupied with clay and makes worms most of the session. One time she asks, "Where's my daddy? He's coming home this afternoon."

T-21. She makes worms of clay. "Worms can't talk . . . Where is my daddy—daddy is dead."

It was months before Leslie expressed her sorrow in an appropriate way. The therapist felt that Leslie's first tears presented another opportunity to bring mother and child closer together.

T-46. Leslie started to cry convulsively. After some preliminary comforting, I took her to her mother who showed her pictures in a favorite story book and pointed to pictures as Leslie, Roy, etc. Then she pointed to a policeman and Leslie identified it as "Daddy—he's in heaven." Back in the playroom she sobbed, "I want my daddy."

The birth of a child to one of her older sisters gave Leslie an opportunity to express feelings of jealousy displaced from her brother.

T-42. Leslie spends the session sucking on the tiny bottle and bathing the baby. She tells me she musn't hit the baby's face hard.

Anal and masturbation problems and castration anxiety were brought out.

T-27. Leslie begins again to play with clay candles and electricity but soon gets into material about "peeing" and "poo-poo". She tells me she touches her poo-poo and then with the rubber genital dolls plays out mother's spanking girl—girl's spanking mother—everybody's being bathed, washed, etc. She says touching poo-poo makes hands ugly and I reassure her that the hands are the same.

T-29. Leslie keeps insisting that girls have penises and over and over again touches them and asks about them as she dresses and undresses the dolls. She keeps confusing herself with the dolls, tying in the boy with Roy and the girl with herself, saying yes she has a vagina—no, she has a penis—yes, she has a vagina, etc. . . . .

Self-mutilation and punishment was a repetitive theme that Leslie expressed through obsessive play. The therapist recognized her fears of destruction and offered her both understanding and protection.

T-23. She makes a Christmas tree of clay and adds candles—she keeps repeating she mustn't put her finger in the flame.

T-26. She knows what she wants to do—she uses clay to make a candle and instructs me to have the little girl put her hand in the flames. I say I won't let the little girl get hurt and I push the doll away each time. Then she makes an electric wire and plays at having a clay doll touch it. . . . little girl mustn't touch the wire . . . little girl mustn't, etc.

T-30. She indicates she wants to put her hands on the electric fan, I say we will not do that . . .

Leslie spent weeks in concern with candles, with electricity, and then in painting and making jails. Throughout, obsessive interest in abstractions, spatial and temporal concepts, were brought out. Leslie painted moons and suns and talked of white, cold snow and of outer space. After the marriage of one of her sisters (about T-100), Leslie drew wedding cakes and brides for several weeks. At the same time, despite the repetitive interests, there were continued gains in relationships both in the Center and within the family. "Did you miss you? . . . . Is mommy afraid you'll catch cold?"

Following Leslie's entrance into kindergarten, she began to ask many questions that indicated she was trying to test reality and trying to get help in understanding man-made rules. She seemed to be saying, "tell me what to do".

T-88. She draws and paints the schoolyard and school-friends. She asks why you have to stay in line and why you have to take turns. She asks questions without waiting for the answers but is not bizarre.

Gleaned from sessions 100 to 130 are the following quotes—showing Leslie's growth and problems.

What happens if mommy calls and you don't like to come? . . . How come nobody likes a tattletale? . . . How old are you when you die—is gray hair old? Why can't we drink pee-pee?"

"I don't want a little brother. What's a worry? Is it a worry when mother comes late? Why aren't real people happy?

Leslie is still in therapy. She has had one session following a planned four-month vacation, including the summer and the first marking period in school. She and her mother brought in her second grade report card, indicating average grades and no particular adjustment problems.[3]

T-131. When I tell Leslie I missed her because I like her, she answers slowly "I think I like you too". She indicates that she still has problems and wants my help in figuring them out. When you were a little girl did you ever have bad thoughts?" She reveals obsessional ideas about kidnappers and says, "It'll be hard to get it out of my head", or "I'll grow out out of it probably?" Leslie reassures me that she doesn't draw kidnappers and then reveals her mother's fear of her expressing these thoughts either at home or at the Center. I try to show Leslie that her mother worries because she loves her.

I suggest that I can tell mother that Leslie is old enough now to have secrets at the Center and that here she can let me know what she is afraid of . . .

Leslie was then able to make two drawings: one of kidnappers for the therapist to see, and one of cats for her mother to see. This appeared to be a conscious effort at repression.

Growth was summed up in various ways. By Leslie, "I told my girl friend that she broke a promise to me and I'll break one to her and she won't like it. Is that what you'd do?" "Remember when we played baby? Remember when I used to draw jails?" And by mother, as Leslie snuggled on her lap and mother hugged her, "You're a big girl now and you can have secrets."

Summary

The first four years of an outpatient treatment program for a severely disturbed child has been described, with particular emphasis on work with the child herself. Despite strong theoretical arguments for the more intensive and all-inclusive program offered in a residential treatment center, an outpatient program within the framework of a child guidance center was attempted. Quotations from the weekly therapy notes reflect the gains that were made by the child.

Two very important features of therapy that may distinguish Leslie's case from that of other children seen at the Center are: (a) treatment has continued long beyond the usual time period, and (b) the child's therapist has remained constant during this four-year period. It can be speculated that this latter circumstance has made it possible for Leslie to feel secure enough to permit the closeness that allowed for symbolic regression, emotional feeding, and eventually the active building up of more useful defenses than withdrawal.

It is important to note that Leslie was able to be regressed in the playroom and at the same time to meet in a limited way the demands of reality at home and in the classroom. Keeping Leslie at home made it possible for Leslie's mother to make some superficial changes in the environment and to respond positively to changes in Leslie herself. The Child Guidance Center has been able to be flexible in planning for more or less intensity in the treatment program depending on Leslie's needs, and is prepared to serve as consultant to the school, and to offer continued help to Leslie, in times of stress, as she grows.

[3]Up to T-130, Leslie s mother has had two social workers. When the second worker left the agency it was decided that mother would continue with Leslie's therapist.

References

1. Bender, Loretta. Schizophrenia in Childhood—Its Recognition, Description and Treatment. *Amer. J. Orthopsychiat.*, 1956, *26*, 499-506.
2. Kaufman, I., Rosenblum, E., Heims, L., and Weller, L. Childhood Schizophrenia—Treatment of Children and Parents. *Amer. J. Orthopsychiat.* 1957, *27*, 683-690.
3. Mahler, Margaret S. On Childhood Psychosis and Schizophrenia. *The Psychoanalytic Study of the Child.* Vol. 7. New York: Internat. Univ. Press, 1952, p. 286-305.
4. Sechehaye, M. A. *Symbolic Realization. A New Method of Psychotherapy Applied to a Case of Schizophrenia.* New York: Internat. Univ. Press, 1951.

ANALYSIS OF INTERPERSONAL RELATIONSHIPS AS AN AID TO UNDERSTANDING FAMILY DYNAMICS: A CASE REPORT[1]

BERNARD GUERNEY, JR. AND LOUISE GUERNEY

Rutgers — The State University

Introduction

There has been an obvious long term trend toward greater emphasis on influencing parents as a means of facilitating treatment of disturbed children. Recently, Ackerman [1] and others [2, 4, 5, 6] have given added impetus to this trend by emphasizing the diagnosis and treatment of the family as a dynamic entity rather than a collection of individuals. Traditional psychological assessment techniques, however, orient clinicians in the latter direction. For purposes of family diagnosis and treatment it would be well, instead, to give greater emphasis to the following considerations. The same conflicts in a given individual will be heightened or minimized, expressed in one way rather than another, elicit one response or another from other persons, *depending* upon what the other individuals in the family are like, how the individual expects the others to behave, and upon how they actually do behave. Significant change in the behavior of one family member automatically will produce some changes in the responses of others in the family toward him. The full understanding of a given family member, therefore, depends upon understanding his interpersonal interactions with the other family members. Fresh theoretical orientations and measurement techniques seem desirable to supplement traditional ones in implementing such considerations.

What appears to be a singularly appropriate approach is available, although there seems to be no rapproachement as yet between its advocates and the major spokesmen for family diagnosis. We refer to Leary's [3] theory and techniques of interpersonal multilevel personality diagnosis. The purpose of this presentation is not to demonstrate what might be done by using all of the techniques, or only the techniques, of the interpersonal diagnostic system. It is intended, rather, to illustrate how the traditional sort of case analysis, presented first below, was complemented by a second analysis employing only a few of the non-quantitative conceptualizations of Leary's system, especially those having to do with "interpersonal reflexes" and the principle of "reciprocal interpersonal relations."

The Case Study

Presenting Problem. Helen was a nine year old white female residing in a lower middle class community, attending the third grade in the neighborhood public school. Her father, Mr. D., was 31 and not yet settled occupationally, having shifted jobs within the area of retailing every year or

[1]This case was seen as part of a Family Group Therapy Research Study being conducted by the authors and supported by a grant from the University Research Council of Rutgers, The State University of New Jersey.

two since returning from the service. The mother, Mrs. D., was only 26 years of age, having married at 16. Helen's early childhood was reported as having been stable and happy, although she had been cared for by the maternal great-grandmother during the day, Mrs. D. having been obliged to work outside the home. Early school adjustment was good in spite of several school changes. In the second grade, while the family was residing in South Carolina, Helen began resisting going to school at home and then resisting again in the classroom. She rebelled to such a degree that the principal would no longer tolerate the disturbance she caused, and insisted that she be enrolled in a private day school for emotionally disturbed children, where she apparently made a good adjustment. At the same time, she was seen privately by a psychologist for about six sessions, with no appreciable results. The services of the psychologist were sought primarily because of the appearance of additional symptoms. Helen had developed an obsession about death, and expressed morbid curiosity, mixed with fears, about childbirth and many aspects of sex.

At the time of referral to the Rutgers Clinic, Helen had recently returned to New Jersey and re-entered a public school. She was repeating her behavior of the previous year, pitifully clinging to her mother and sobbing, exhibiting violent displays of temper including biting, striking, refusal to move, abusive language, screaming, and last minute illnesses to prevent herself from being sent to school. Consultation with the school psychologist revealed that there was apparently little concrete about the school situation itself to account for this behavior. When she was settled in the classroom for the day, Helen's adjustment and academic performance were good. The parents reported that Helen had become completely unmanageable at home, refusing to abide by any limits, and preventing (by means of the behavior described above) the parents from going out together without her. In addition, fears of death, growing up, and sex were expressed with ever-increasing frequency.

Test Behavior. The test battery consisted of the WISC, Goodenough, Bender-Gestalt, Michigan Picture Test, and puppet play observation. The parents were interviewed, and each filled out an Interpersonal Check List for self, spouse, and child. Both parents appeared intelligent, spoke freely and openly about the problem, and seemed well-motivated to do something about it. Mrs. D. was the more verbal of the two, particularly with regard to describing her daughter's "faults". Mr. D. tended to tone down his wife's statements and expressed himself with more control.

Helen, herself, was a thin, wiry child with an intense, nervous manner and a flippant, self-conscious way of responding to attention. She related easily, and quickly expressed feelings of love and possessiveness toward the examiner, primarily through notes that she wrote. Just as quickly, she expressed irritation and impatience with the tests or the examiner when the tests became difficult or conformance was demanded. There was extreme testing of limits along with efforts to control the situation with threats, pouting, and withdrawal of affection. In general, however, it was obvious that Helen wished to succeed, enjoyed doing so, and was self-conscious about failure.

Summary of Traditional Evaluation. Helen was of average intelligence and showed no signs of neurological impairment. Dependency gratifications seemed to have been minimal. Infantile narcissistic needs still exerted a strong influence on Helen's perceptions and behavior. Aggressive impulses were strong, particularly toward her mother, whom Helen regarded as rejecting. Conflict between a great need for dependency gratification from the mother and the hostility resulted in marked ambivalence toward women, inadequate feminine identification, and sexual role confusion. Competition with her mother, based on strong Oedipal conflict, made the dependency-hostility conflict still harder for Helen to resolve. Father figures were seen as ineffective, but not rejecting.

Helen's ego was weak, and she was quite unsuccessful in gaining adequate satisfaction for her needs. Superego development was uneven, and emotional controls were poor. Impulses were apt to break through in extreme and inappropriate ways. Reality testing was not, however, seriously deficient. Helen's self concept was not so much negative as it was tentative and ambivalent. She vacillated between feeling worthwhile and adequate, even powerful, and very unloved and inadequate. Naturally, a great deal of anxiety was generated by these conflicts and uncertainties. Helen's major defenses against these anxieties seemed to be denial and displacement.

The presenting problem seemed to be explained by these findings. The overdependence, combined with hostility toward the mother, prompted her inability to attend school without rebellion. This was a perfect vehicle for satisfying the two opposing drives. The sex problem arose from her confusion regarding her sexual role, and the fears of death and separation again from the conflict of hostility versus dependence. Fears of growing up seemed to arise from all of these dynamics working together to make the future most insecure and forbidding in prospect. Her strengths were her adequate intelligence, the existence of some positive self percepts, and, most important, the fact that she continued to desire to win affection from others and continued to strive, albeit in ineffective ways, toward this end. Since the roots of this problem seemed to go back to the earliest stages of psychosexual development, the prognosis was guarded.

Summary of Analysis of Interpersonal Family Dynamics. In the clinician's eyes, both parents seemed weak and ineffectual in their interpersonal mechanisms as evidenced in the interview, by the ease with which Helen manipulated them, and by the father's apparent vocational instability and lack of direction. This was quite a different picture from their own self-perceptions.

They did not see one another or themselves as weak, but as fitting in with cultural expectations of adequate authoritativeness and responsibility.

Consciously, Mrs. D. saw herself as sympathetic and forgiving, responsible and generous. She saw herself as having little in common with her daughter, whom she saw as calculating, thinking only of herself, self-satisfied, boastful, and rebellious. The father, on the other hand, though he recognized rebelliousness in his daughter, emphasized this much less than the mother, and was more prone to see good qualities such as leadership and even "making a good impression on others" as descriptive of her. His perceptions of Helen were rather similar to his consciously reported perceptions of himself.

Helen apparently saw the mother as alternating between tender, sympathetic, and soft-hearted behavior on the one hand, and aggressive, rejecting, punitive, and hypercritical behavior on the other. She saw the father as being capable of aggressive outbursts, but as predominantly cooperative, affectionate, and understanding. Helen regarded herself as being the victim of her mother's rejection and the "darling" of her father. Her self-perceptions, though, seemed to be more a reflection of her mother's behavior toward her: extreme vacillation between love and hate. (Helen wrote frequent notes to her parents which were almost invariably signed, "Love or hate, Helen.")

Docile-dependent attributes in all the family members were conspicuous by their absence at the level of consciously reported perceptions. However, a pattern of dependency on an unconscious level was present in all members of the family. Dependency was particularly strong and important in Helen's case. She behaviorally expressed dependency to a marked degree, but mainly when it could be expressed in a rebellious, aggressive context; that is, when it ran counter to the immediate wishes of the parents, as when they wished to go out for an evening, or when they wanted Helen to be independent and fulfill her social and cultural obligations. Thus, her dependency was rarely expressed in such a way as to draw the reciprocal nurturant and warm responses from the parents, particularly the mother. The reasons for the lack of direct and appropriate expression of dependency seemed to be Helen's hostility and competitiveness toward her mother, which also demanded expression, and the fear that her mother's response to a direct bid for nurturance would be overprotection and domination. The closest she could come to "pulling" the nurturant responses, unmixed with hostility, from the mother was through feigned and/or psychosomatic illness. Through this behavior she could continue to be the one in control while receiving sympathy and warmth. Moreover, she could terminate the situation when protection threatened to become overprotection, simply by becoming well again.

A rejecting reaction to Helen's rebellion-cloaked dependent behavior was made all the more likely and powerful because of Mrs. D.'s image of what she should be like as a mother. Her ideal was to be a strong, nurturant, and protective mother, which involved the wish and expectation for reciprocal behavior of obedience and respect, trust and acceptance, from Helen. She wished Helen were dependent upon her for emotional satisfaction and guidance. Helen's deviations from this idealized picture produced extreme vacillation of feeling and behavior in the mother. She alternated between self-effacing and self-derogatory behavior at one moment, when she felt that she had been deficient in living up to this image, and resentment and aggressive behavior the next moment, when she placed the blame for destroying this ideal on Helen.

The pattern, of course, was sometimes initiated from the other side as well, with the mother the first to show rejecting or punitive behavior toward Helen, calling forth a response of counter-rejection and rebellion, which, in turn, reinforced the mother's negative response. It was often some demonstration of healthy independence on Helen's part that called forth Mrs. D.'s initial critical response. It was assumed that the mother's distorted interpersonal reflex patterns and conflicting interpersonal motivations greatly contributed to the development of the inadequate pattern of interpersonal reflexes later shown by Helen in this area. But the counter-reflexes developed by Helen made it much harder than it originally would have been for Mrs. D. to express warmth without domination, and therefore harder to break the vicious cycle.

Docility and dependence were important characteristics of Mr. D. While he did not perceive these aspects in himself, he probably expressed them most directly and openly of the D.'s. Faced with one of Helen's demanding moods, Mr. D. would usually leave the house rather than take an active part in attempting to control Helen or join his wife in aggressive harangues. Mr. D. occasionally did control Helen with angry, punitive outbursts, but most of the time he was virtually helpless in preventing her from taking advantage of him. This was apparently the case not only because of his docility, but because of dependence upon her for gratification of his concept of himself as a kind, generous, understanding, sympathetic person in general, and parent in particular. Dependency upon his wife was more pronounced. He depended on her not only in the emotional area, but in more practical matters as well, e.g., budgeting, fulfilling social obligations, and handling problems with neighbors, in-laws, etc. While Mrs. D. and Helen were usually the antagonists in the family dissensions, Mr. D. seemed to contribute to the high level of conflict if only because of his general ineffectiveness in the interpersonal area and because his dependency on both would not permit him to take sides. With no clear allegiance to either one, he remained the target for and object of both mother and daughter's respective efforts to gain the dominant position.

Helen's interpersonal reflexes were the most limited, automatic, and overdeveloped, and by virtue of this fact it was she who dictated the nature of most of the family's interaction. The ease with which Helen could step into the position of dominance and control over her mother, and the

frequency of the mother's descent to the level of using child-like recriminations and arguments with her daughter lent an element of reality to Helen's Oedipal fantasies. Helen's confusion in sexual role and her masculine strivings were seen as emerging from repeated reinforcement for expressing authoritarian and dominating interpersonal reflexes, from hostility engendered toward the mother in the blind struggles mentioned above, from the inability to "pull" warmth and sympathy from her mother, and from the identification the father felt for Helen, which gave rise to a reciprocal identification on her part.

Therapy. The interpersonal dynamics seemed to be such as to present an interesting case for joint mother-daughter therapy sessions. It was felt that this might present the quickest route to a breakup of the interpersonal misunderstandings and vicious cycles. Joint therapy sessions including the father were not possible because of his employment. The mother and daughter were seen together by a single therapist, followed immediately by separate therapeutic sessions for each, again with the same therapist. Discussion of therapeutic techniques would not be appropriate in this paper, but it does seem appropriate to give some form of closure to the reader by reporting that it appeared to the therapist that the joint interviews provided a unique opportunity for mother and daughter to express their deeper feelings about each other directly to each other in a situation allowing for therapeutic intervention. In addition, the joint sessions seemed to serve as a catalytic force in promoting the working out of the feelings and conflicts in the individual sessions. The interpersonal themes mentioned above were very dramatically displayed and worked through in the therapy. The case was closed after a total of 22 forty-minute individual sessions for each, and 17 twenty-minute joint sessions. The therapist's opinion, test results, and the family's reports at that time indicated that, in addition to a cessation of the presenting problems, healthy patterns of communication and interpersonal behavior had largely replaced the former pathological ones. There was similar evidence that Helen's problems of sexual role identification had been satisfactorily resolved. Whatever interpersonal and intrapersonal conflicts remained, no longer seemed outside of the family's capacities to work through without further professional help.

Summary

In this paper the potency of traditional psychological orientations and techniques was questioned in the light of recent attempts to understand and treat families as a unit, as a group in psychological interaction. The system of interpersonal diagnosis as developed by Leary was viewed as a system of great promise for this purpose. The case of a nine year old girl who refused to go to school, had fears of death, fears of growing up, and was confused in sexual role, was presented in traditional fashion, and as a family study utilizing some concepts from Leary's interpersonal system, in order to illustrate the utility of these concepts for understanding family dynamics in action.

References

1. Ackerman, N. W. *The Psychodynamics of Family Life.* New York: Basic Books, 1958.
2. Goolishian, H. A. *A Brief Psychotherapy Program for Disturbed Adolescents.* Paper read at Amer. Orthopsychiat. Assn. Chicago, 1960.
3. Leary, T. *Interpersonal Diagnosis of Personality.* New York: Ronald Press, 1957.
4. Rapaport, Rhona. The Family and Psychiatric Treatment: a conceptual approach. *Psychiat.*, 1960, *23*, 53-62.
5. Schuster, F. P., Jr. Summary Description of Multiple Impact Psychotherapy. *Texas Rep. on Biol. Med.*, 1959, *17*, 426-430.
6. Smith, B. K. *A Family Grows.* Austin, Texas: The Hogg Foundation for Mental Health, The Univ. of Texas, 1959.

INDIVIDUAL THERAPY WITH A "DEFECTIVE DELINQUENT"

ERWIN FRIEDMAN

Children's Psychiatric Center, Eatontown, N. J.

Introduction

The elusive concept that the so-called "defective delinquent" is not amenable to rehabilitation is still a widely held opinion among institutional administrators. Westwell[11], for example, quotes one superintendent as recommending that "the defective delinquent" should be triple locked up and permanently "taken out of circulation." Others have expressed the view that psychotherapeutic intervention is unlikely to eliminate delinquent behavior. Thus, Lurie *et al*[6] concluded that "prognosis with regard to cure of behavior difficulty is uniformly poor." In recent years, Sarason[8], Stacey[9] and Thorne[10] have expressed a more positive opinion regarding psychotherapeutic endeavors with mental defectives in general and defective delinquents in particular. While environmental therapy has been stressed and documented in some English studies[3, 7, 12, 13], and group therapy results have been published in this country[1, 2, 5], there seems to be a paucity of reports of individual psychotherapy. This paper offers a case study with a defective delinquent who was rehabilitated through intensive individual psychotherapy. The subject was selected by the staff of a large Colony for male defectives as belonging to the group of "incorrigible" defective delinquents. The procedure of selection and the results of a large scale counseling and psychotherapy program have been described elsewhere.[4] One to three therapeutic sessions of 20-50 minutes each were held weekly for 18 months except for vacation.

Case Report

Peter is a twenty-four year old single man. Both parents were born in Russia, had no formal education and came to this country after World War I. Father died in 1942. Mother died in 1954. The marriage was reported as a stormy one. His mother was considered a good housekeeper but completely unsuccessful in controlling her son; her attitude was one of both rejection and over-indulgence. Two siblings are college graduates. Pre-natal and developmental history is sketchy and not reliable. Peter was an unwanted child. His birth was normal and he had generally good health except for pneumonia at one year of age. School attendance began at the normal age but he was excluded almost immediately. Since entire family life was upset by Peter's presence in the home, school was reapproached but again resulted in exclusion because he was "unmanageable, inattentive, deliberately disobedient and resorted to attention-getting in the classroom situation." Following his second exclusion from school, it was decided that Peter should be tried in a foster home placement. He was placed in 1942 but ran away and it proved physically impossible to contain him. In October, 1943, Peter entered a residential treatment center for disturbed delinquents and truants in New York State. While there he continually invited punishment from his fellow inmates. He nagged, provoked and used all possible devices to direct attention toward himself. His aggressiveness was manifested by clownish, buffoonlike bantering and persistent demands and questions. He bit his nails severely, had a voracious appetite and wet the bed several times a night. Psychotherapy was initiated and immediately resulted in a marked change in social manners, improved school adjustment and less severe nail biting and enuresis. Unfortunately, the therapist had to leave and Peter could not adjust to his new therapist. He became explosive and destructive, and on one occasion jumped on her,wrapping his legs around her body and kissing her. The staff felt that the institution could not cope with Peter's destructive and disruptive behavior and recommended transfer to a psychiatric hospital. The mother did not agree and took the boy home in December, 1946.

At home Peter refused to obey his mother and tried in every way to annoy her. He would run outside and ring the doorbell until she became extremely upset. Since the mother was unable to cope with him, admission to a Residential Treatment Center in New Jersey was arranged. The diagnostic impression given after his admission in April 1950 was "Disorder of personality due to epidemic encephalitis." The exhibited behavior difficulties were in the form of constant talking, interfering with the activities of other children, hyperactivity, provocative behavior, etc. He also on numerous occasions felt the breasts of women supervisors and of strange women visiting the Center. Peter responded poorly to the various therapies instituted for him and on reaching the Center's age limit was released in February 1950. A subsequent evaluation at the New Jersey Diagnostic Center resulted in a diagnosis of "psychopathic personality with sexual deviations, compulsive type, possibly associated with organic encephalopathy". Peter was transferred to a reformatory. In his four years there, he seemed unable to fit in any program and was regarded as "our most abnormal boy". Constant violation of the regulations resulted in segregation and punishments. He was paroled April 1954 but because of voyeuristic activities, drinking, etc., parole was revoked and Peter was recommitted four months later to the State Colony for Defectives. During the first two years his behavior was similar to that experienced before, with some further decline noted. Psychological testing resulted in IQ ratings of 72, 51, 57, for the Verbal, Performance and Full Scale of the Wechsler-Bellevue. The Wide Range Achievement Test showed the following grade equivalents: Reading 5.2, Spelling 4.8 and Arithmetic 3.1. Until 1957 Peter spent most of his time in the detention cottage, repeatedly sent there for breaking institutional rules, and for homosexual and aggressive acting-out behavior.

The Course of Psychotherapy

After reviewing Peter's case during the Spring 1957, the writer established a therapeutic relationship with the patient. The first six months of individual psychotherapy produced little noticeable effect. Peter continued his behavior pattern and made frequent trips to the detention cottage where therapy was continued. He verbalized feelings at that time that he was a bad boy and nothing would make him change his ways, that this was the way he wanted to be, and he didn't care whether his behavior would result in a complete and eternal confinement to the detention cottage or some jail. However, at the end of the six month period, it seemed that the very permissive atmosphere which the therapist created for Peter and the fact that the therapist did not lose interest at the time of his difficulties with institutional administration had finally made an impression on him. He seemed to acquire some social skills and ideas and made a promise that he would at least try to change some of his behavior patterns, would try to keep himself out of detention, and would take an assignment as a worker. Previous to this time Peter was unable to hold any work assignments; he had been tried in many training situations throughout the institution and had constantly failed. During these first six months, Peter's behavior in the therapy sessions could be characterized as extremely polite and subdued, as he tried to please the therapist. He never failed to thank the therapist for his efforts to help him, was extremely courteous, and tried to do small favors to gain the therapist's confidence and approval.

During the second six month period, Peter's behavior changed considerably. He was very proud of his new ability to keep out of mischief and the absence of any disciplinary action. He counted every day from his last confinement in detention. Now, however, the therapist became an object of Peter's extreme hostility and aggression. He came to the sessions accusing the therapist of all kinds of things, telling him that everyone in the Colony was trying to help him except the therapist. He questioned the value of the hour and the therapist's ability to help him. He brought something new to every session, attempting to discover if this would make the therapist feel bad or perhaps reject him. He asked repeatedly what the therapist would do if he would attack his wife (who at that time was a teacher in the Colony

school.) Quite often he would remain outside the door of the therapist, not coming in but also making sure that the therapist would not become occupied with another boy or another activity.

After the three week Summer vacation period, Peter greeted the therapist with friendliness. He expressed his sincere feelings of having missed the therapist, talked about his own inadequacy and inability to face the problems of the world, and of how he really feared going out into a strange and hostile community. This was the first time he was able to express hostility toward his deceased mother who had rejected him, placed him in institutions and never really cared about him. Similar feelings were directed against other members of his family who would come once-in-a-while to visit him or send small gifts, but were not really interested. This period was therapeutically very fruitful. He brought his dreams to the session and gained quite considerable insight in their meaning. He was also able to discuss every-day matters and his behavior toward peers and supervisors. His judgment and insight in these situations appeared quite mature and developed more and more as therapy continued. The cottage parents' reports from this period forward showed a marked improvement in his behavior and working habits. He was amenable to counseling and, probably for the first time in his life, accepted a job and performed very well.

A schedule was set up for Peter to go home for weekends and gradually for a week. It was felt that this might lead to a good relationship with the family, which could eventually result in a job placement with supervision from home. It proved a very trying period for the brothers. Peter turned all his aggression and hostility toward them, especially at the time of the older brother's marriage. He created numerous disturbances during his home visits, running away, annoying everybody, stealing money from the younger brother, etc. His behavior was discussed in the therapy sessions and interpreted as an attempt to make the brothers reject him, which in turn, would give him justification for his delinquent behavior. Peter finally gained insight and after a long period of difficulties a satisfactory relationship developed between him and his brothers.

A formal psychological test was administered in the Spring of 1958 and a marked improvement in all areas was evident. On the WAIS, the Verbal IQ increased from 72 to 76, Performance IQ from 51 to 66, and Full Scale IQ from 57 to 70. On the Wide Range Achievement Test, his grade level for Reading went up from 5.2 to 8th grade and Spelling from 4.8 to 5.1 grade level. The projective tests showed that Peter was more comfortable with his home environment and less aggressive. The concluding phase of therapy was not too different from the previous one. Peter was again able to bring his feelings and his problems to the sessions and discuss them in a frank and orderly manner. His behavior improved even more and, as noted above, the family contact and family relationship grew stronger and he felt at ease in the new setting. He went home more frequently, enjoying the life of the community and the relationship with his family. One very important change occurred during this last period. Peter was always a ravenous eater and was markedly overweight. Beginning approximately at the end of Autumn, he lost close to forty pounds, which seemed a good indication of lessened orality, of gaining inner security and balance. It was the unanimous opinion of the Colony Staff that Peter had made an outstanding adjustment; and his behavior on work assignment was excellent and not once had it been necessary to confine him to detention.

Peter was placed in the community approximately two years ago with the understanding that he would continue receiving supportive therapy during his community life and work. Unfortunately, this did not materialize. During this period Peter did not become involved in any delinquent acts; there were no reports of sexual misbehavior and, as far as his over-all behavior was concerned, his adjustment was considered fair. His work record, however, was only marginal. He still preferred to live on his brother's earnings, and was not willing to make a wholehearted effort to take a full time job. There were some suggestions that the writer should continue psychotherapy with Peter until a complete readjustment to the outside world was achieved but, for technical reasons, this was not feasible.

Conclusion

Individual psychotherapeutic intervention in the case of a young "defective delinquent" was described and it was shown that, contrary to other reports in the literature, prognosis can be favorable. It is suggested that institutions and agencies dealing with borderline or mildly retarded delinquents should make extensive use of psychotherapeutic techniques. The importance of intensive counseling after release from the institution was emphasized. Instead of supporting the widely prevailing attitude of pessimism and custody, the case presented may offer some hope for rehabilitating defective delinquents, helping them to become more productive citizens and reducing the burden of society.

References

1. Cotzin, M. Group psychotherapy with mentally defective problem boys. *Amer. J. ment. Def.*, 1948, *53*, 268-283.
2. Fisher, H. A. and Wolfson, J. N. Group therapy of mental defectives. *Amer. J. ment. Def.*, 1953, *57*, 463-480.
3. Foale, M. An approach to stabilization of male juvenile mental defective delinquents. *Amer. J. ment. Def.*, 1952, *57*, 116-122.
4. Friedman, E. Defective delinquents reconsidered. Read at the 84th annual meeting of the Amer. Assoc. ment. Def., May 1960.
5. Kaldeck, R. Group psychotherapy with mentally defective adolescents and adults. *Int. J. group Psychother.*, 1958, *8*, 185-192.
6. Lurie, L. A., et al. The defective delinquent, a definition and a prognosis. *Orthopsychiat.*, 1944, *14*, 95-103.
7. O'Connor, N. and Yonge, K. A. Methods of evaluating the group psychotherapy of unstable defective delinquents. *J. genet. Psychol.*, 1955, *81*, 89-101.
8. Sarason, S. B. Individual psychotherapy with mentally defective individuals. *Amer. J. ment. Def.*, 1952, *56*, 803-805.
9. Stacey, C. L. and DeMartino, M. F. (Eds.) *Counseling and psychotherapy with the mentally retarded.* Glencoe, Ill.: The Free Press, 1956.
10. Thorne, F. C. Counseling and psychotherapy with mental defectives, *Amer. J. ment. Def.*, 1948, *52*, 263-271.
11. Westwell, Arthur E. The defective delinquent. *Amer. J. ment. Def.*, 1951, *56*, 283-289.
12. Weber, H. The borderline defective delinquent. *Brit. J. Delinq.*, 1953, *3*, 173-184.
13. Yonge, K. A. and O'Connor, N. Measurable effects of group psychotherapy with defective delinquents. *J. Ment. Sci.*, 1954, *100*, 944-952.

PROBLEMS OF PSYCHOLOGICAL DIFFERENTIAL DIAGNOSIS ASSOCIATED WITH HYPOTHYROIDISM: A CASE STUDY

REBECCA SNYDER

Veterans Administration Hospital
East Orange, New Jersey

Introduction

The necessity for recognizing that the psychopathology of patients suffering from organic disturbances may simulate very closely disturbances of a functional nature is knowledge which is sometimes easily lost sight of when florid manifestations of a thought disorder are evidenced and leave no uncertainty about the presence of a psychosis. Failure to differentiate organic psychoses from functional psychoses before the occurrence of irreversible organic damage deprives the patient of the potential benefits of medical intervention which such a differentiation might make possible. Asher[1] points out the truism that amelioration of the condition depends upon its recognition and describes the psychoses of hypothyroidism as conditions which go unrecognized by physicians not infrequently and are misdiagnosed as schizophrenia. The disorientation, ideas of reference, and delusions of persecution commonly found in such conditions contribute to this confusion[1, 2, 3]. Bleuler[2] was aware of the difficulty and described some "rare cases of hypothyroidism scarcely

distinguishable from schizophrenia". The Clinical Society of London, too, in 1888 made known its recognition of the psychotic aspects of myxedema. Asher[1] quotes from this report as follows: "Delusions and hallucinations occur in nearly half of the cases, mainly where the disease is advanced". More recently Asher[1], Cranswick[2] Gregory[3], and Zondek and Wolfsohn[5] are among those describing "schizophreniform" psychoses associated with thyroid hypofunctioning and clearing subsequent to thyroid medications. Asher[1] believes that this condition exists much more frequently than is recognized. He found that 9 out of 14 psychotic patients referrred to him within a period of 4 years recovered after thyroid medication although none of them had been diagnosed by the referring physician as having hypothyroid conditions. On the basis of this experience he concluded that there must be many others which go undiagnosed because of the lack of specificity of the resultant psychoses.

The problem of differentiating psychoses associated with thyroid hypofunctioning from functional psychoses is no less confusing for the psychologist since he too is confronted with a lack of specificity in the nature of the resultant psychoses. Swartz and Ferguson[4] presented a case report of a patient whose psychopathology had been, prior to testing, considered by her therapist to be functional while the interpretation of the test findings suggested a possible organic dysfunction with the subsequent identification of hypothyroidism as "a major organic factor". The case being presented here is that of a patient with an established diagnosis of hypothyroidism. A psychological evaluation was requested because of "bizarre behavior" which he was exhibiting. His physician raised the question of a possible functional psychosis. At the time of the referral request, the report of the patient's previous hospitalizations had not yet been received so that neither the referring physician nor the examiner was aware of the patient's having formerly been diagnosed as suffering from paranoid schizophrenia.

Case Report

The patient was a 44 year old bachelor whose medical history indicated a condition of hypothyroidism and hypopituitarism, apparently of longstanding duration. Paranoid ideation was florid and had contributed to his having formerly been diagnosed as having paranoid schizophrenia. Characteristics of his psychological functioning suggested to this examiner the possibility of an organic psychosis in which paranoid features were prominent. The occurrence of both an organic psychosis with prominent paranoid features and the condition of hypothyroidism suggested to the examiner the possibility of a relationship between the two, a "schizophreniform" psychosis associated with the hypothyroidism. Since the patient died 16 months subsequent to the psychological examination without having shown any remission of his psychosis despite thyroid medication, proof of such a relationship was not established. Autopsy findings, however, did reveal "atrophy of the adrenals, thyroid and pituitary glands" so that failure of thyroid medication to bring about the remission of psychosis could have resulted from the presence of irreversible organic dysfunction. We have no way of knowing whether earlier thyroid medication might have alleviated his psychotic symptoms. The purpose of presenting this case is to share with others the problems of differential diagnosis which arose and to focus attention upon a condition which may well be escaping our attention and being categorized as schizophrenia.

Psychological Examination. The patient was seen on two different occasions within a three day time interval and given the following tests: Wechsler Adult Intelligence Scale, Bender-Gestalt and Rorschach. Suspiciousness was clearly apparent in the qualitative nature of the patient's initial interaction with the examiner. It was as if he were enveloped by it and withdrawn into himself so as to make as little contact as possible with potentially harmful people. He seemed almost mummified. Unlike the typical schizophrenic, however, he seemed gradually to lose some of his mistrust and became more outgoing and animated. He even tried to be helpful and at one point made a rather pathetic attempt at being humorous. Paranoid ideations, however, sporadically erupted in very dramatic form. He asked whether the person in the next office who could be seen through a window between offices was taking his picture, and on questioning indicated that he knew that was one technique that the FBI used when trying to find out about anyone. He added that he did not believe this was being done to him. He described his life on the ward as one in which he was being "humiliated" by the other patients. Most striking of all, however, was his drastic reaction to Card VI of the Rorschach. This evoked in him what he considered "evil" thoughts. This he projected upon the cards by calling them "evil" and indicating that the examiner was trying to do something evil to him by showing him these cards. He asked whether the card had "anything to do with a woman"; then he put it down as if he were being contaminated through contact with it. He asked the examiner what she was "driving at" and added when asked his thoughts about this, "If it's what I'm thinking, you'd better stop."

Sexual confusion was apparent as he revealed his close identification with a sister who has "bladder trouble" and numerous bodily complaints for which her physician could find no organic basis. He sent her to a psychologist who indicated that experiences of the past were now influencing her. His defensiveness was temporarily gone as he asked the examiner whether she thought this might also be true of him. Feelings of guilt and thoughts of evil soon intruded into his thinking again and he became almost incoherent as he communicated in fragmented phrases and telegraphic style the content of his disturbing thoughts. They concerned his having looked at such "evil" pictures in the past, in smokers and men's club rooms and feeling guilty about it. He spoke in agitated manner of the "evil" pictures he had seen and of the need "to confess to the Priest whatever is in your thoughts." "I went to Confession and don't want no more of these . . .", he stated with vehemence. He believed that there was "something evil within" him which no one could help, neither the physicians nor anyone else. Paranoid ideation, fragmented thinking, loosening of associations, flattened affect, the ready use of projection, and erratic intellectual functioning with success on some items more difficult than those he failed, indicated a paranoid psychosis, resembling paranoid schizophrenia.

Qualitative rather than quantitative aspects of this patient's functioning suggested to this examiner the probability of an organic rather than a functional psychosis. The following features of his functioning seemed atypical of the schizophrenic and more characteristic of the organic: dissatisfaction with his own functioning on concrete tasks and the acceptance of the limitations as being within him, feelings of inadequacy, perseveration, feelings of perplexity and confusion, misuse of words, pattern cohesion, concretistic thinking, a subjective feeling of having slowed down in his responses, the spontaneous moments of helpfulness, involvement in the tasks, attempts at improving his performance, and the gradual emergence of clear thinking from the confused state in which direct questions frequently threw him. It was as if he needed more time than usual to fully grasp the essence of that which he was being asked.

His obtained Full Scale IQ of 93 was still within the normal range, Verbal IQ, 96 and Performance IQ, 91. The erratic nature of the patient's intellectual functioning was reflected in the intra- as well as intertest scatter. His performance on the Digit Symbol subtest was equivalent to an IQ of 79. This occurred despite his ready understanding of the nature of the task and was a reflection only of his inability to work quickly. There were no errors, simply an insufficient number of blocks filled. His highest score was obtained on Block Designs (equivalent to IQ 112). Pattern cohesiveness and perseverative tendencies prevented his obtaining an even higher score. His approach to both of these subtests of the WAIS impressed this examiner as organic.

The patient's handling of the Bender-Gestalt drawings showed no distortions of designs but some perseveration. He was concerned with the accuracy of his reproductions and tried to do well. Dissatisfaction with his own designs was expressed as were requests for further structuring from the examiner.

The Rorschach was not completed since the patient refused to continue after seeing Card VI. Up until that time form level was good and concretistic thinking conspicuous.

The patient's psychological functioning thus revealed florid examples of disordered thinking such as are characteristic of a paranoid psychosis. Features suggestive of a possible organic impairment were also present so that the diagnosis of an organic psychosis seemed most applicable to the total picture. The organic manifestations were, however, much less dramatic and conspicuous clinically than the paranoid ones so that the patient's behavior in many ways so closely resembled that of a paranoid schizophrenic as to raise the question of which was primary or whether the two conditions coexisted with no causal relationship between the two. A careful and accurate history about the nature of the patient's personality changes could have been helpful in attempting to resolve this problem, but there was no reliable source from which to obtain this. The history which he gave indicated a thyroid condition of long standing and the possibility of irreversible changes having occurred was recognized.

Summary

The problem of differential diagnosis between organic psychoses associated with hypothyroidism and schizophrenia is focused upon in this case presentation of a patient with paranoid psychosis and a severe hypothyroid condition. It is suggested that the organic psychosis with paranoid features such as has been described might have been associated with the hypothyroidism. Definitive evidence for this was not however established since thyroid medication failed to bring about a remission of the psychosis.

References

1. Asher, R. Myxoedematous Madness. *Brit. M. J.*, 1949, *2*, 555.
2. Cranswick, E. H. Tracer Iodine Studies on Thyroid Activity and Thyroid Responsiveness in Schizophrenia. *Amer. J. Psychiat.*, 1955, *112*, 170-178.
3. Gregory, T. Mental Disorder Associated with Thyroid Dysfunction. *Canad. Med. Assoc. J.*, 1956, *75*, 489-492.
4. Swartz, M. and Ferguson, E. D. Interview and Test Data in a Diagnostic Problem: A Case Report. *J. clin. Psychol.*, 1959, *15*, 124-127.
5. Zondek, H. and Wolfsohn, G. Myxoedema and Psychosis. *Lancet*, 1944, *2*, 438.

COMPARISON OF LEITER, WAIS AND STANFORD-BINET IQ'S IN RETARDATES

JOHANNES C. BRENGELMANN AND JOSEPH T. KENNY
State Colony, Woodbine, N. J. *Lower Penns Neck Public School, N. J.*

Problem

Considerable IQ discrepancies in groups of retardates have been obtained between the Leiter International Performance Scale (LIPS), the Stanford Binet (SB) and the Wechsler Verbal (WAIS-V) and Performance (WAIS-P) scales [1, 3, 5]. In some experiments, Performance IQ was superior, in others Verbal IQ. Variation in intelligence level was thought to contribute significantly to these discrepancies. This study investigated differences between the above tests over a wide IQ range. The following sources of variation are considered: test difficulty in terms of overall differentiation between tests; nature of test in terms of a Performance/Verbal test dimension; intelligence level in terms of five grades maximally varying as to IQ; and, finally, kind of intelligence level in terms of three level criteria varying along a Performance/Verbal dimension.

Method

Subjects. Seventy-five retardates, who completed all tests, were drawn from the resident helper male population of Woodbine State Colony. The main selection criterion served to produce greatest possible IQ variation.

Tests. The tests selected are the LIPS, a pure performance scale; the SB (Form L), a mixed performance/verbal scale; the WAIS V and P subscales which are dealt with separately because of the differential effects between verbal and performance tests. The sample means, with standard deviations in brackets, are as follows: LIPS = 47.40 (14.14), SB = 50.59 (11.55), WAIS-V = 60.03 (7.92), WAIS-P = 61.69 (14.30), and WAIS Full Scale = 58.47 (10.56). They increase in the order LIPS - SB - WAIS, as expected from the literature.

The correlations between the four test scores are as follows: LIPS *vs.* SB = 0.77, LIPS *vs.* WAIS-V = 0.59, LIPS *vs.* WAIS-P = 0.75, SB *vs.* WAIS-V = 0.79, SB *vs.* WAIS-P = 0.78, and WAIS-V *vs.* WAIS-P = 0.71. With one exception (LIPS *vs.* WAIS-V), these coefficients are fairly uniform.

For cross-validation of the Wechsler Performance/Verbal discrepancy effects, described in a separate paper [2], a score of WAIS-P minus V (P-V) was derived. The present mean is 1.67 (SD = 10.52).

Chronological age. Age varied considerably about the mean of 36.91 years, SD being 11.88. Correlations between this factor and the test scores are as follows: LIPS = −0.12, SB = 0.00, WAIS-V = 0.13 and WAIS-P = 0.26. Only the latter coefficient is significant at the 5% level, indicating that age plays no significant role in the present analysis.

Procedure. The tests were administered in a balanced order to control for sequence effects. The Binet and Wechsler tests were scored in the conventional manner. The scoring procedure for the LIPS was as suggested in a personal letter by Dr. R. G. Leiter: MA multiplied by 100 and divided by a constant of 156 with 5 IQ points added (for *Ss* of more than 13 years). This procedure differs somewhat from that described in the manual [4].

The criteria for the Performance/Verbal dimension were the LIPS, the rank order combined WAIS-V: LIPS and the WAIS-V score. The first and last score differ most with regard to both the Performance/Verbal dichotomy and intercorrelation. Variation in intelligence level was effected by dividing the sample into non-overlapping quintiles on the basis of each of the three criteria of the Performance/Verbal dimension (Table 1, quintiles L to H).

Results

The main findings are given in Table 1. Test means are shown as a function of the five intelligence levels and three intelligence criteria. The slope of the individual test gradients, as determined by the quintile means, is steepest for LIPS and WAIS-P, the two performance scores and flattest for WAIS-V, a verbal score. The gradient slope is further dependent on the kind of IQ level criterion. It increases in the order Verbal—combined Verbal/Performance—Performance criterion.

Table 1. Test Means and *t*-Ratios Between Means as a Function of Five Intelligence Levels (Sample Quintiles)

Intelligence Level	L	ML	M	MH	H
			Mean IQs		
WAIS – V criterion					
LIPS	35.7	39.0	49.5	52.6	60.3
SB	39.0	42.2	51.0	55.9	64.9
WAIS – V	49.5	54.9	59.4	65.2	71.1
WAIS – P	49.0	53.3	63.1	65.4	77.6
WAIS – V: LIPS criterion					
LIPS	29.3	40.1	48.5	53.7	65.4
SB	36.1	45.1	50.8	56.0	66.7
WAIS – V	50.9	55.4	59.5	64.9	69.3
WAIS – P	44.4	55.9	62.6	68.9	76.7
LIPS criterion					
LIPS	27.3	39.7	47.3	55.9	66.8
SB	38.0	45.5	51.5	55.4	62.5
WAIS – V	54.2	57.0	59.1	64.3	65.6
WAIS – P	47.1	55.5	62.4	65.7	77.8
			t Ratios		
WAIS – V criterion					
LIPS *vs.* SB	1.1	1.0	0.5	1.1	1.5
LIPS *vs.* WAIS – V	4.5***	5.1***	3.2**	4.1***	3.5***
LIPS *vs.* WAIS – P	4.3***	4.6***	4.4***	4.2***	5.6***
SB *vs.* WAIS – V	3.4***	4.1***	2.1*	3.0**	2.0*
SB *vs.* WAIS – P	3.2**	3.6***	3.9***	3.1**	4.1***
WAIS V – *vs.* WAIS – P	0.2	0.5	1.2	0.1	2.1*
WAIS – V: LIPS criterion					
LIPS *vs.* SB	2.7**	2.0	0.9	0.9	0.5
LIPS *vs.* WAIS – V	8.5***	6.0***	4.3***	4.4***	1.5
LIPS *vs.* WAIS – P	5.9***	6.2***	5.5***	5.9***	4.5***
SB *vs.* WAIS – V	5.8***	4.0***	3.4***	3.5***	1.0
SB *vs.* WAIS – P	3.2**	4.2***	4.6***	5.0***	3.9***
WAIS – V *vs.* WAIS – P	2.6*	0.2	1.2	1.6	2.9**
LIPS criterion					
LIPS *vs.* SB	3.8***	2.1*	1.5	0.2	1.5
LIPS *vs.* WAIS – V	10.4***	6.2***	4.3***	3.0**	0.4
LIPS *vs.* WAIS – P	7.1***	5.7***	4.8***	3.5***	4.0***
SB *vs.* WAIS – V	5.8***	4.2***	2.7**	3.2**	1.1
SB *vs.* WAIS – P	3.3**	3.6***	3.9***	3.7***	5.5***
WAIS – V *vs.* WAIS – P	2.6*	0.6	1.2	0.5	4.4***

L = low, M = medium, H = high intelligence level. All means based on N = 15. DF = 280 (Remainder of Analysis of Variance).

Significance level: * = 5%, ** = 1%, *** = 0.1%

Significance of these results was assessed by means of analysis of variance. F-ratios for the WAIS-V criterion were as follows: Tests = 51.5, level of intelligence = 84.0, and T × L = 0.5; for the combined WAIS-V: LIPS criterion: T = 74.8, L = 148.2, and T × L = 2.5; for the LIPS criterion: T = 63.7, L = 106.6, and T × L = 5.1. The DF are: T = 3, L = 4, T × L = 12, and Remainder = 280. The *F*-ratios for tests are significant at the 0.1% level. The test by level interaction (variation in gradient slope) proves significant and significantly dependent on the level criterion.

Along the criterion dimension, this interaction increases from insignificance (Verbal criterion) to a 0.1% level of significance (Performance criterion).

We may now proceed to test the significance of more specific effects using *t*-ratios between all tests for IQ levels as shown in Table 1. The respective Mean Squares of the Remainder served as variance estimates. They are 71.4 for the WAIS-V criterion, 49.2 for WAIS-V: LIPS, and 57.8 for LIPS.

Point 1. Significant differentiation is obtained between any combination of two tests and at most points of the analysis. More than two-thirds of the *t*-ratios are significant at the 5% level, the average ratio being 3.3. The significance of differentiation decreases steadily from low to high IQ level, as follows: L = 4.35, M L = 3.51, M = 2.98, M H = 2.82 and H = 2.78 (*t*-ratio means), *i.e.*, means differ most widely at the lowest IQ level. The mean difference between the LIPS and WAIS-V, for example, amounts to 26.9 IQ points (Table 1, L level of LIPS criterion).

Point 2. The difference in gradient slope produces a significant crossing-over effect between the two Wechsler Scales. The Verbal IQ is significantly higher than the Performance IQ at the L level, while the opposite applies for the H level. This confirms the result of a separate investigation[2]. More pronounced crossing-over effects would, of course, have been obtained if sample means between tests had been identical, particularly in the case of the LIPS/WAIS – V interaction.

Point 3. For the Verbal criterion, average differentiation between tests is very similar for all IQ levels, and the otherwise characteristic Performance/Verbal effects are almost neutralized. As to the remaining two level criteria, effects specific to the Performance/Verbal dimension become more pronounced and differentiation, particularly at low levels, increases. This result throws light on the relative efficiency of intelligence tests in discrimination between individuals.

Wechsler Performance-Verbal discrepancy. Previously obtained relationships between P – V and intelligence[2] are practically duplicated. P – V correlates 0.40 with LIPS, 0.49 with SB and 0.63 with WAIS (Full Scale). All coefficients are significant at the 0.1% level (N = 75). Relative or absolute superiority of Performance, therefore, correlates positively and significantly with intelligence, regardless of the intelligence test used.

Chronological age, again, correlates positively with P – V, the coefficient of 0.23 just reaching the 5% level of significance. Age and WAIS Full Scale, from which the P – V score is derived, are themselves correlated to a degree of 0.26. Partialling out intelligence, the correlation between age and P – V is 0.09, not significant. However, WAIS Full Scale correlates 0.61 with P – V after partialling out the effects of age. This suggests that, with intelligence held constant, age does not correlate significantly with P – V.

Discussion

Some IQ discrepancies are apparently due to overall task difficulty rather than kind of test items used. A comparison between the LIPS and WAIS-P scales shows, for example, that differences are fairly uniform for all conditions represented in Table 1. Differences of this kind may be easily eliminated by adjusting scoring procedures.

Other discrepancies appear dependent on the kind of test, whether Verbal or Performance. Differentiation between individuals is best for the Performance scores, as indicated by their greater steepness in gradient slope. Even the use of a Verbal score as criterion of IQ level (WAIS-V) does not lead to superior individual differentiation of Verbal over Performance scores. This may be due to item heterogeneity of the WAIS-V or the SB. Some items may be more strongly accumulative with regard to Verbal and others with regard to Performance intelligence. The discussed P – V discrepancies could, of course, be removed by analysing both verbal and performance items against the same intelligence criterion. However, in the interest of more specific measurement it is suggested that such differences be maximized rather than concealed or, in other words, that specific rather than general aspects of intelligence be emphasized. Efficient prediction of other criteria of adjustment and/or personality may greatly depend on specialization and purification of test scores.

The comparison of pure Performance and Verbal tests may possibly lead to new insight into the problem of trainability and/or educability of the retardate. It may

be contended that, as Performance falls increasingly short of Verbal capacity, trainability decreases. The crossing-over point, where superiority of Performance or Verbal IQ is established, may perhaps serve as a cutting point for the prediction of trainability. To investigate this problem, more refined Performance and particularly Verbal measures of intelligence are required. This includes the construction of tests capable of discriminating more sensitively between moderately and severely retarded individuals. The LIPS appears a promising step into this direction.

Summary

The Leiter International Performance Scale, the Stanford-Binet (Form L) and the Wechsler Verbal and Performance Scales were administered to 75 male retardates. Differences between tests were analyzed as a function of intelligence level, there being three level criteria: a Verbal, a combined Verbal/Performance and a Performance criterion. Considerable differences were obtained between all tests. The degree of differentiation was complexly determined by test difficulty, kind of test (whether Verbal or Performance), intelligence level and kind of level criterion. Difficulty was assumed to simply increase overall differentiation. The Verbal/Performance test dimension interacted significantly with intelligence level, Verbal IQs being superior at low levels and inferior at high levels. The degree of this interaction, again, is dependent on level criterion. With a verbal criterion, the interaction was insignificant, while a Performance criterion produced greatest and highly significant interaction.

References

1. Beck, H. S. and Lam, R. L. Use of the WISC in predicting organicity. *J. clin. Psychol.*, 1955, *11*, 154-158.
2. Brengelmann, J. C. Brain injury, intelligence level and Verbal/Performance discrepancy in retardates. (Submitted for Publication).
3. Gallagher, J. J., Benoit, E. P. and Boyd, H. F. Measures of intelligence in brain damaged children. *J. clin. Psychol.*, 1956, *12*, 69-72.
4. Leiter, R. G. *The Leiter International Performance Scale.* Chicago: Stoelting, 1948 (Rev.)
5. Miller, M. B. Psychometric and clinical studies in mental deficiency. 1954-1959: A selective review and critique. *Amer. J. ment. Defic.*, 1960, *65*, 182-193.

ANXIETY AS "EFFECT OF UNCERTAINTY": AN EXPERIMENT ILLUMINATING THE OA SUBTEST OF THE WAIS[1]

EILEEN S. LANFELD[2] AND D. R. SAUNDERS

Educational Testing Service, Princeton, N. J.

Introduction

A factor analysis[6] of split-half scores from the Wechsler Adult Intelligence Scale (WAIS) did not yield any factor clearly identifiable with Object Assembly (OA), although eleven factors were extracted and all the other subtests except Comprehension exhibited reliable unique variance. However, a subsequent study[7] focused on the items of the Picture Completion (PC) subtest did produce one factor clearly related to OA. Therefore, it seems likely that Object Assembly confounds at least two variables: a "performance" component held in common with Block Design (BD) and, as Rapaport[4] has suggested, an anxiety component held in common with certain PC items. Further verification of an anxiety component in OA may be sought by devising other measures that should correlate with it especially

[1]This research was supported by the Society for the Investigation of Human Ecology.

[2]Present address: Israel Institute for Applied Social Research, Jerusalem, Israel.

well. Most recently, this component has been found in a broader context of tests including the Afrikaans Wechsler [3], but still has not been shown to involve any non-"performance" tests.

We have hypothesized specifically [5] that the potentially anxiety-provoking feature of the OA task is that the subject is not given any model to copy, so that he must manipulate and organize the pieces of each puzzle in order to learn what it makes. If this ignorance proves bothersome to the subject, we may call him anxious. If this "effect of uncertainty" hypothesis is correct, then the similar *verbal* task embodied in the Sentence Arrangement Test to be described below should correlate with OA and PC but not with BD.

Procedure and Results

The Sentence Arrangement Test (SA) and the Picture Arrangement subtest of the WAIS are very similar in their mechanics of administration. However, instead of containing a picture, each of the cards laid before the subject in SA has from one to three words printed on it in large capital letters.[3] The cards for each item are presented in a standard random order, and the subject's task is to arrange them in sequence so the words "say something." The time required to reach a grammatically correct arrangement provides the basis for scoring. In a pilot experiment, using six adult subjects, we observed that taking logarithms (to the base 10) resulted in detection of a linear relationship between length of sentence and time required for solution.

Experiment 1. Hypothesis: SA correlates with over-all ability as well as do any of the regular WAIS subtests.

Fifteen items, ranging in length from eight to nineteen words (four to ten cards), were administered to twenty-one new subjects, aged eighteen to twenty-five, each of whom had previously taken the WAIS. The subjects' normal levels (NL) on the WAIS[4] ranged from 9 through 18.

The product-moment correlation coefficient between NL and total log-time for this sample was $-.78$. This result is highly significant statistically, indicates a marked relationship between SA performance and intelligence as measured by the WAIS, and is in line with the hypothesis.

Analysis of variance of the 15 by 21 table of log-time scores showed that sentence characteristics other than length can play a significant role in performance on SA. By plotting item difficulty against sentence length, we found that two of the fifteen items were abnormally difficult. These items were " / He had / the knowledge / that what / we love we / may also / despise. /" and " / About all / her problems, / Carol, at / sixteen, was / not in / the least / worried. /" We interpreted the latter as an artifact caused by the complex punctuation, and the former as possibly due to the potentially threatening content of the item.

Experiment 2. Hypothesis: SA's correlations with WAIS subtests, particularly PC, BD, and OA, are as stated in the introduction.

The total log-time score on SA was correlated separately with each of the WAIS subtests, using the data of experiment 1. In order to remove the effect of general intelligence from these correlations, we first converted the subtest scores into deviations from NL for each subject, and these deviations were correlated with SA by the product-moment formula. The resulting values are shown in Table 1. I, OA, and PC are the only subtests significantly related to SA when holding NL constant. Both OA and PC have been shown elsewhere [3, 7] to involve the "effect of uncertainty" factor, which we deliberately incorporated in Sentence Arrangement.

[3] Examples: / For the / tenth time / he got / up and / looked out / of the / window. // In the morning I / bought three / tickets for / the two- / o'clock bus / to town./ A complete list of the SA items employed may be found in [1].

[4] NL equals the sum of weighted test scores (WTS) for I, C, D, A, S, PA, PC, and BD, minus WTS for OA, all divided by 7. For further explanation see [4, 8].

TABLE 1. CORRELATIONS OF SA WITH WAIS (N = 21)

(I – NL) vs –SA:	.53**	(PA – NL) vs –SA:	.00
(C – NL) vs –SA:	.00	(PC – NL) vs –SA:	.43*
(D – NL) vs –SA:	.08	(BD – NL) vs –SA:	.24
(A – NL) vs –SA:	.35	(OA – NL) vs –SA:	.45*
(S – NL) vs –SA:	.14	(DS – NL) vs –SA:	.03

**Significant beyond the 1% level (one-tailed test).
*Significant beyond the 5% level (one-tailed test).

The high correlation with Information is appropriate for a verbal test and serves to replace the high correlation typically encountered between OA and BD.

Experiment 3. Hypothesis: Performance on SA may be adversely influenced by employing sentences incorporating indirect psychological threat.

Thirteen items were administered to nineteen new subjects, aged eighteen to twenty-five. The items included eight well-behaved items from experiment 1, the two deviant items, and three new items: / If / you / believed that / he / might shoot / our mother / I / would give in / right away. // Since / money / is strong / and / God / is even / more powerful, / there / is nothing / to fear. // Really, / when / we are / alone / together / I cannot / take time / for sleep; / what / should I / do? / The latter three items were newly written in a deliberate effort to incorporate a degree of threat that would be apparent only to subjects attending to the meaning and implications of the sentence as a whole.

The 8 by 19 table of log-time scores for the presumedly normal items was subjected to an analysis of variance summarized in Table 2. (The analyses of variance

TABLE 2. ANALYSIS OF VARIANCE FOR EXPERIMENT 3

Effect	Variance	O f	Var/Of	F-ratio
Grand Mean	151.700	1	151.700	4311.34***
Persons	6.273	18	.348	11.27***
Sentence Length (linear regression)	12.073	1	12.073	335.62***
Sentences, not Length (deviation of means from line)	1.176	6	.196	6.69***
Residual ("Error")	3.896	126	.0309	(1.00)
Totals	175.118	152		

***Significant beyond the 0.1% level.

were performed following McNemar's model for "analysis of bivariate correlation". (2, pp. 270 ff.)) Since the error variance for this experiment does not differ significantly from that of the first experiment, the results are probably comparable. Some reduction was achieved in variance due to sentence effects other than length in comparison with the variance due to sentence length itself, but the reduction is not significant statistically. The effects of persons, sentence length, and sentences not-length were all statistically significant in both analyses of variance.

In this experiment, if difficulty is plotted as a function of sentence length for the eight "normal" items and the five experimental items are compared with the resulting regression line, each of the five is found to deviate significantly from the line at better than the 1% level, by one-tailed test and in the expected direction of greater difficulty. We concluded, therefore, that performance on SA is influenced not only by sentence length but also by sentence content, that a major effect of content may be interpreted in terms of threat, and that items can be written that deliberately incorporate this threat. The threat may act to increase the "effect of uncertainty" in the items involved, and thus lengthen the time required for at least some subjects to arrange them.

Summary

We have described some preliminary steps in the construction of a Sentence Arrangement Test, designed for individual administration in conjunction with the WAIS. High mean performance on an SA item is associated with the use of a short sentence, the avoidance of internal punctuation as a cue to the solution, and the avoidance of psychological threat in the content of the solution. High SA performance by an individual subject is dependent on general intelligence, a factor of verbal information, and a factor of freedom from the "effect of uncertainty". (The latter is regarded as equivalent to "situational anxiety".) It appears to be possible to construct and administer a reliable form of SA that is valid for these factors.

References

1. Lanfield, E. S., and Saunders, D. R. Some determiners of performance on an experimental Sentence Arrangement Test. *Res. Memorandum 59-12.* Princeton, N. J.: Educational Testing Service, 1959.
2. McNemar, Quinn. *Psychological Statistics.* (2nd Ed.) New York: Wiley, 1955.
3. Mundy-Castle, A. C., and Saunders, D. R. Intelligence and brain rhythms in a socially isolated community. (In preparation)
4. Rapaport, David. *Diagnostic Psychological Testing, Volume I.* Chicago: Year Book Publishers, 1945.
5. Saunders, D. R. An outline of Gittinger's personality theory as applied to the Wechsler: I. The subtests considered separately. *Res. Memorandum 59-3.* Princeton, N. J.: Educational Testing Service, 1959.
6. Saunders, D. R. On the dimensionality of the WAIS battery for two groups of normal males. *Psychol. Rep.*, 1959, *5*, 529-541.
7. Saunders, D. R. A factor analysis of the Picture Completion items of the WAIS. *J. clin. Psychol.*, 1960, *16*, 146-149.
8. Saunders, D. R. Digit Span and Alpha Frequency: A cross-validation. *J. clin. Psychol.*, 1961, *17*, 165-167.

AN EVALUATION OF RORSCHACH RELIABILITY THROUGH THE ALTERNATE-RESPONSE METHOD

HARRY STEIN

New Jersey Diagnostic Center, Menlo Park

Problem

The traditional method of determining Rorschach reliability has been by the split-half method and the usual method of splitting the cards has been by comparing the responses on the odd cards to the responses on the even cards[1, 5, 6]. The major assumption of the odd *vs.* even split is that this division produces an equality of stimulus values[2]. To the extent that discrepancies exist in the stimulus value of odd *vs.* even cards, a minor assumption is involved, *i.e.*, that all responses to a card are determined directly by the card. If a person gives more than one response to a card, he must, so to speak, eradicate the first response and return to the card for the stimulation of his next response. This assumption is necessary for the following reason. If the individual's second response to any card is determined by his previous response rather than the card stimulus, we may see a "run" of responses which will exaggerate any discrepancies in stimulus values between the odd and even cards. This would result in a lower split-half reliability than could be indicated by an alternate response method of determining reliability.

A concrete illustration of the problem might be helpful. In a test employed clinically by the author, the patient is requested to write down as many different words as he can in three minutes. An individual writes down the following: "Mother, father, sister, brother, table, chair, lamp, ashtray, cat, rat, lion, tiger, India, China, Spain, Italy". (This pattern of responding is not uncommonly given by a hospital-

ized patient population.) Let us suppose each group of 4 words were responses to one stimulating idea, or, relating it to the Rorschach, each group of 4 responses was given to a Rorschach card. Thus, by the traditional method of splitting the cards, we would be comparing 'mother, father, sister, brother' with 'cat, rat, lion, tiger', and 'table, chair, lamp, ashtray' with 'India, China, Spain, Italy'. Obviously this makes no sense since we would be comparing family members with animals and office equipment with countries and the net results would be high reliability for R but zero reliability for content scores. However, the alternate response method would compare mother to father, sister to brother, cat to rat, lion to tiger, etc. This method does make sense out of the subject's responses, if one employs a method of scoring the responses such that the similarity in content is revealed.

A further illustration might clarify both the problem and the hypothesis tested in this study. If one were to find a number of Rorschach records with 20 responses, of which 2 were given to each card, would there be a greater similarity between responses given to the same card or would the similarity be greater across cards ? The hypothesis was that a comparison of reliabilities obtained for 31 Rorschach scores by the alternate response method would be greater than obtained by the split-half method.

Method

Since it would be difficult, time consuming and perhaps even impossible to find a sufficient number of 20 response records, each of which has 2 responses per card, the author used a sample of 53 individuals with varying degrees of physical disability, described elsewhere[4], where the average number of responses to the Rorschach was 19.6. The Klopfer and Kelley[2] method of administration was employed with some minor modifications[4].

Because the literature[3] has indicated a "succession" effect in the pattern of W, D, d, Dd, a method was devised for equalizing this factor across the split consisting of grouping the cards with respect to the number of responses given. Thus,

Figure 1. Method and Illustration of Dividing a 20 Response Rorschach Record to Calculate Alternate Response Reliability[1]

Cards	I	II	V	VII	III	IV	VI	VIII	IX	X
No. Responses Per Card	2	4	2	2	1	1	1	3	1	3
Starting Position	A	B	A	B	B	A	B	A	B	A

Placement of Each Response

	A	B
Card I	1	2
Card II		1
	2	3
	4	
Card III		1
Card IV	1	

etc.

[1]What the author was attempting to avoid with this method of dividing the responses was the case of an even number of responses to each card all taking the form of W, D, d, Dd. In this case, a simple alternation would place all the W and d responses on one side of the split and the D and Dd responses on the other side of the split.

all cards with an even number of responses were alternately started on different sides of the split starting with A, while those cards with an odd number of responses were alternated starting with B. An example of the division of responses is given in Figure 1.

Results and Discussion

A comparison of the reliability correlations obtained by the split-half and alternate response methods is given in Table 1. Instead of an improvement in the reliability correlations, there was a non-significant decrease from the average reliability of .246 for the split-half method to .216 for the alternate response method. The hypothesis that the reliabilities of Rorschach scores would be improved by the alternate response method must be rejected.

Table 1. Comparison of Alternate Response and Split-Half Reliabilities with Usual Method of Rorschach Administration.*

Score Per Cent	Split-Half Reliability	Alternate Response Reliability
W%	.459	.628
D%	.286	.507
d%	.555	.244
Dd%	.289	.348
S%	.023	.221
Location Average**	.334	.410
M%	.058	.179
FM%	.283	.257
m%	.273	.166
F%	.354	.527
k%	−.102	−.109
Fc%	.247	−.079
cF, c%	.214	−.021
FK%	.322	.104
K, KF%	.108	.151
C′%	.286	.518
FC%	.318	.295
C, CF%	.230	.193
Determinant Average**	.218	.191
H%	−.012	.104
A%	.472	.439
Aobj%	.058	.041
Anat%	.075	.079
Obj%	.094	.047
Nat%	.546	.685
Plant%	.156	.253
Geo%	.484	.247
Art%	.445	.311
Arch%	.433	.110
Cloud%	.535	−.054
Blood%	.000	.000
Fire%	−.116	−.116
Emb%	−.058	−.062
Content Average**	.240	.161
Grand Average**	.246	.216

* The reliabilities given below have *not* been corrected by the Spearman-Brown Prophecy Formula.
** The averages were determined through *z* transformations.

An examination of the table and cogitation about the development of the hypothesis led to the realization that, in the administration of the Rorschach, the author had observed "runs" or sequence effects for such scores as Dd and F. Certain obsessive persons on some cards would give a long run of Dd's whereas others would give long runs of F responses. It was this experience which had led through various stages to the formulation of the hypothesis. It would appear that a resolution of the clinical observation of "runs" and its relationship to reliability is more complex than

indicated by the slight reduction in average reliability from the split-half to the alternate response method. The data in Table 1 suggest that, for certain types of responses, there is a facilitating effect in that the likelihood of a similar response is enhanced, whereas for other types of responses, the emission of a similar response is inhibited.

Let us examine the data for inhibition. If a person gives a response indicative of anxiety, this disquieting emotional state would predispose him to return to the card and search out a response as different as possible than the one he had given. Examining the group of shading responses (k%, Fc%, cF,c%, FK%, and K,KF%), a sharp drop in reliability occurs from the split-half method of determination to the alternate response method. The average for the 5 shading scores by the split-half method is .161, whereas it is only .009 by the alternate response method. This interpretation is supported by the sharp drop in the reliability of Cloud responses for the alternate response method (from .535 to −.054). On the other hand, there is no support in the Anatomy responses, which remains the same for both methods of determining reliability.

There is a general decline in reliability for all content scores from an average of .240 by the split-half method to .161 by the alternate response method. It has often been observed by the author, and others, that patients frequently make self-critical comments about their inability to diversify their content on the Rorschach. Such comments as, "All I can see are animals" and "What, another bat?" indicate a general desire to vary their content. The precipitous drop in reliability from the split-half to alternate response method for such content categories as Geography, Architecture and Cloud indicate that one such response tends to inhibit a similar succeeding response.

The average reliability of Location scores rises from .334 as measured by the split-half method to .410 when determined by the alternate response method. This would suggest that the utilization of a W, D, etc., type of response, facilitates an identical type of response. Similarly, when an individual gives a Form response, he tends more strongly to give an identical response immediately following than he is likely to give a Form response to another blot as a new stimulus situation.

Several factors must be kept in mind regarding the above analysis of the data. Since the initial hypothesis was not supported, the interpretations of the results are in the nature of new hypotheses requiring experimental confirmation. Also, the specific procedure employed in dividing the responses so as to avoid the "succession" effect may have affected the reliabilities obtained with the alternate response method. Ignoring the "succession" effect, which was minimal in the records, and taking a simple alternation may be a better procedure to employ. Although the reliabilities of individual scores are quite variable[4], the data in Table 1 strongly suggest a facilitating "runs" effect for certain of the Rorschach scores, including the Location scores and F%. An inhibitory effect was observed in the Content and Shading scores. Should these observations be confirmed, we can accept these Location and F% alternate response reliabilities as being "truer" measures of reliability than those obtained by the split-half method.

For purposes of experimental design it was necessary to assume that Rorschach responses, other than the first one, are either determined by the stimulus of the card or the *preceding response*. The response, of course, may be determined by both in varying degrees at different times during the test. For this condition, which is probably the most accurate description of emission of Rorschach responses, neither the split-half nor alternate response method of determining reliability will give the full measure of Rorschach reliability.

Summary

On the basis of observations of "runs" of certain Rorschach scores, it was hypothesized that reliability of Rorschach scores would be greater when measured by the alternate response method than by the usual split-half method. This hy-

pothesis was not supported. However, the group of Location scores and F% yielded higher reliabilities by the alternate response method. An inhibitory effect on the successive emission of Shading and Content scores was seen in the lowered reliabilities by the alternate response method than by the split-half method. If these observations are confirmed in additional studies they would suggest greater reliabillty for Rorschach scores than is presently presumed.

References

1. Hertz, M. R. The reliability of the Rorschach ink-blot test. *J. appl. Psychol.*, 1943, *18*, 461-477.
2. Hertz, M. R. Current problems in Rorschach theory and technique. *J. proj. Tech.*, 1951, *15*, 307-338.
3. Klopfer, B. and Kelley, D. M. *The Rorschach method.* Yonkers-on-Hudson: World Book Company, 1946.
4. Stein, H. Rotation and reliability of the Rorschach. *J. proj. Tech.*, 1960, *24*, 171-181.
5. Thornton, G. R. and Guilford, J. P. The reliability and meaning of Erlebnistypus scores in the Rorschach test. *J. abnorm. soc. Psychol.*, 1936, *31*, 324-330.
6. Wirt, R. D. and McReynolds, P. The reliability of Rorschach number of responses. *J. proj. Tech.*, 1953, *17*, 493-494.

HOMOSEXUALITY AND SOME MMPI MEASURES OF MASCULINITY-FEMININITY[1]

BERNARD S. AARONSON AND HOWARD R. GRUMPELT

N. J. Bureau of Research in Neurology and Psychiatry
Box 1000, Princeton

8 South Wells Street
Kingston, Pennsylvania

Problem

A number of recent studies[1, 6, 9] have suggested that the personality differences between men and women comprise more than a single dimension. The persistent stereotype that homosexuals are in personality more like individuals of the opposite sex seems as a result to require re-evaluation. In addition, the varying proportion of homosexual outlet noted by Kinsey, *et al.*[4, 5], and the delineation by Hathaway[2] of a number of different kinds of homosexuals, raise further question as to the validity of this concept. If the stereotype is true, then personality measures which differentiate men from women should also differentiate homosexuals from heterosexuals, even though the correlation between these measures is low.

Procedures

To evaluate this, a sample of profiles obtained on the Minnesota Multiphasic Personality Inventory (MMPI) from 25 males with histories of exclusive or nearly exclusive homosexuality was gathered. Fourteen came from the New Jersey State Hospital at Trenton, seven from a VA hospital in Illinois, and four from a state hospital in Minnesota.[2] These were matched with 25 male patients at the New Jersey State Hospital with no history of any homosexual activity or preoccupation at all. Matching was on the basis of age and the type of patient population from which the individual cases were drawn. That is, those from a criminal unit were matched with patients from a criminal unit, those from an acute service with those from an acute service. The mean age of the homosexual group was 28.0, and of the heterosexual group, 28.3, which is not significantly different.

[1]The basic data were gathered while both authors were on the staff of the New Jersey State Hospital at Trenton.

[2]The writers are indebted to Dr. Robert E. Jones, Donald R. Peterson, and Mr. Roland M. Peek for their assistance in gathering these data.

The T score on the Mf scale (Mfs), the rank of this scale in the profile (Mfr) consisting of the ten clinical scales set forth in the test manual[3], and the score on the masculinity-femininity index (Mfi) based on a relationship among the scales Hy, Pa, Hs, and Pt[1], was determined for each profile. These measures were intercorrelated to determine their relationship in each group. The extent to which each measure segregated homosexuals was tested by combining the two groups, eliminating those cases that fell at the median, and comparing the resultant distributions by chi square. The extent to which high or low scores on any of the measures was associated with homosexuality and heterosexuality was determined by comparing the mean scores on these measures in the two groups by *t*-test.

Results

In the homosexual group, Mfs correlates −.60 with Mfr, significant beyond the .001 level, and −.28 with Mfi. In the heterosexual group, Mfs correlates −.58 with Mfr, significant beyond the .001 level, and .25 with Mfi, Mfr correlates .31 with Mfi among the homosexuals, and .19 among the heterosexuals. Only Mfs and Mfr show any appreciable relationship, but it should be noted that these are two different ways of scoring the same scale. The negative sign to these correlations simply reflects the ordering of T scores from low to high as opposed to the ranking procedure in which a rank of one is assigned to the highest score.

The chi square values obtained when the ability of each measure to differentiate homosexuals from heterosexuals was compared were 14.54 for Mfi, 17.94 for Mfs, and 18.42 for Mfr. All of these are significant beyond the .001 level. All of the measures seem to differentiate homosexuals from heterosexuals.

The optional cutting score for Mfi in differentiating these groups is at an index value of 46. Only nine homosexuals fall at or above this, and nine heterosexuals below. The cutting score for Mfs is a T score of 60. Only four homosexuals fall below this, and six heterosexuals at or above. The cutting score for Mfr is a rank of 6. Only one homosexual falls below this and eight heterosexuals above. The optimal combination of these measures as indicators when applied to the present sample is to diagnose homosexuality when any two of the three deviate in a homosexual direction with respect to their cutting scores. This results in misclassifying one homosexual and four heterosexuals. A criterion of three misclassifies twelve homosexuals and one heterosexual, and a criterion of one, one homosexual and eighteen heterosexuals.

The means and standard deviations on the three measures in both groups are shown in Table 1. While the differences on Mfi are in a predicted direction, the

Table 1. Mean Scores and Standard Deviations on the Mf Index, mf Scale Score, and Mf Score Rank in the Homosexual and Heterosexual Groups

	Homosexual		Heterosexual	
Measure	Mean	S.D.	Mean	S.D.
MFI	45.4	14.1	51.6	13.6
MFS	70.4	10.1	55.2	10.2
MFR	3.3	1.8	7.0	2.6

variance in both groups seems large and the *t*-ratio is 1.55, which is not significant. The *t* value obtained for the Mfs means is 5.24, significant beyond the .001 level. Mfr showed much more variability among heterosexuals than homosexuals, as manifested by an *F* value of 2.06, significant beyond the .05 level. The *t*-ratio, calculated with a formula for samples of unequal variance, was 5.78, significant beyond the .001 level. Although all three measures differentiate homosexuals from heterosexuals, and can be used in concert to make such differentiations, marked differences in score associated with membership in these groups appear only with Mfs and Mfr. In-

spection of the data on Mfi shows that a small subset of homosexuals seems to get high, or very "masculine" scores on this index although the larger part of this group falls within the "feminine" range.

Discussion

These data suggest that while male homosexuals in general seem to be more "feminine", a subset seems to exist who show more masculine behaviors. This is in accord with the earlier findings of Terman and Miles[8], whose scale was derived on the basis of essentially the same comparison as Mfi. The measures which differentiate homosexuals from heterosexuals best are originally derived from a comparison of homosexuals with heterosexuals. The overall data raise the question of whether more differences exist between homosexual and heterosexual males than might be anticipated from the action of a dimension of masculinity-femininity considered alone. While a personality orientation similar to that of the opposite sex seems conducive to the expression of homosexual behavior, it does not seem of equal relevance in all instances. The relationship may be more correlative than causal. Conversely, since these groups do behave differently on measures related to masculine-feminine personality differences, the non-unitary nature of these differences is again suggested.

The cutting scores for the several measures as given here probably require revision in normal clinical use. The sample employed was carefully selected and the problem is different than selecting homosexuals from a random clinical population. Similarly, the 90% correct classification with a criterion of any two of the three measures pointing to homosexuality is undoubtedly contaminated by having been applied back to the sample from which the cutting scores were derived. Its significance is in showing that taking a combination of any two of these measures indicating homosexuality as the basis for diagnostic judgment provides the most accuracy.

Summary

Application of three MMPI measures of masculinity-femininity to samples of 25 homosexual and 25 heterosexual male patients shows that while all three differentiate homosexuals from heterosexuals, a subset of homosexuals exists who seem very masculine in orientation. While feminine orientation seems related to homosexual behavior, it does not seem equally relevant in all instances. These data suggest the heterogeneity of any postulated dimensions of masculinity-femininity.

References

1. Aaronson, B. S. A comparison of two MMPI measures of masculinity-femininity. *J. clin. Psychol.*, 1959, *15*, 48-50.
2. Hathaway, S. R. Scales 5 (Masculinity-Femininity), 6 (Paranoia), and 8 (Schizophrenia). In Welsh, G. S. and Dahlstrom, W. G. *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956, 104-111.
3. Hathaway, S. R. and McKinley, J. C. *The Minnesota Multiphasic Personality Inventory Manual (Rev. ed.)* New York: Psychological Corporation, 1951.
4. Kinsey, A. C., Pomeroy, W. B. and Martin, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.
5. Kinsey, A. C., Pomeroy, W. B., Martin, C. E. and Gebhard, P. H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953.
6. Reed, M. R. The masculinity-femininity dimension in normal and psychotic subjects. *J. abnorm. soc. Psychol.*, 1957, *55*, 289-294.
7. Shepler, B. F. A comparison of masculinity-femininity measures. *J. consult. Psychol.*, 1951, *15*, 484-496.
8. Terman, L. M. and Miles, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.

EYSENCK'S "STRATIFICATION" THEORY AND THE QUESTIONNAIRE METHOD OF MEASURING PERSONALITY[1]

C. M. FRANKS, E. A. HOLDEN AND M. PHILLIPS

New Jersey Neuro-Psychiatric Institute, Princeton

Introduction

There is considerable disagreement concerning the equivalence of the information obtained from objective behavior ratings or check lists, direct observation and self-rating questionnaires. For example, Cattell and his associates [8] assume that an almost perfect correspondence between behavior ratings and self-rating questionnaires has been established, whereas Becker [1] arrives at quite a different conclusion after a critical review of the evidence. The similarities and differences between the data obtained from these kinds of personality measuring devices may well be dependent both upon the general nature of the personality factor it is hoped to measure and the population being studied.

It is possible that both techniques yield meaningful information about a subject and that the data obtained from either can be consistently integrated into the same personality model without excessive modification of the theory in question. There is clearly a need for further investigation of this problem and the parameters pertinent to the discrepancies between self- and behavior ratings in various populations. Meanwhile, the absence of such knowledge can present a considerable hazard to those who need to have an objective measure of observable personality behavior, especially if their concern is with the experimental investigation of theories which attempt to relate observable personality differences to various attributes of the central nervous system. Of such theories, that of Eysenck is undoubtedly the most clearly and provocatively formulated and it is, therefore, hardly surprising that it has received the most attention.

Eysenck [2] integrates constitutional factors, learning theory, and personality theory into a stratified system in which there are four main levels, all conceived of as being causally related to each other in hierarchical order. Level 1, the most fundamental, is physiologically determined; level 2 is that of observable experimental laboratory phenomena such as eyeblink conditioning or figural after-effects; level 3 is that of objectively observable primary personality traits such as sociability and activity; level 4 is that of observable attitudinal phenomena such as ethnocentricism and hedonism. The direct concern here is not with the merits or demerits of this theory *per se* but with one particular methodological aspect associated with its investigation. Therefore, while it is fully recognized that the above brief description does scant justice to Eysenck's carefully developed arguments, it is not necessary here to elaborate the theory in any great detail. It is sufficient to reaffirm that the fundamental level, governing a wide variety of behavior at different levels, refers to some constitutional, but at present largely hypothetical, property of the nervous system of the individual.

As Eysenck [2] points out, causally the progression is from level 1 to level 4. Historically, however, levels 3 and 4 were investigated first and independently, and then linked up experimentally with level 2 and hence, by inference, with level 1. The delineation of level 3 began with the description of two orthogonal personality dimensions, namely introversion-extraversion and neuroticism, these concepts being themselves defined in terms of observed relationships between various primary traits such as sociability, excitability, and impulsiveness. These dimensions are now conveniently measured by self-rating questionnaires such as the Maudsley Personality Inventory (MPI) [7], although originally a variety of other measures were used.

[1]The authors wish to acknowledge the help of Miss Elinor Phillips in collecting and scoring part of the basic data. They are deeply indebted to Dr. David R. Sanders for his invaluable guidance in the treatment of the data and the provision of computational facilities. The project was supported jointly by the National Institute of Mental Health under Research Grant M2720, the Society for the Investigation of Human Ecology and the New Jersey Neuro-Psychiatric Institute.

In terms of Eysenck's "stratification" theory, level 2 phenomena, themselves dependent upon the more fundamental processes of level 1, in their turn partly determine level 3 phenomena. For example, the hypothetical central processes of *excitation* and *inhibition* at level 1 are supposed to determine in part whether a subject is good or poor at forming conditioned responses in the laboratory. And a person who is good at forming laboratory conditioned responses is, therefore, presumed to be good at forming conditioned responses outside the laboratory—he is presumably responsive to the stimuli in his environment and to the social climate, sensitive to new people and situations and cautiously discriminating in his observable behavior. Similarly, the poor laboratory conditioner is presumed to be insensitive to new situations and the social climate around him and indiscriminating in his observable behavior. These characteristics may not unreasonably be described as the *observable* aspects of introversion and extraversion respectively. There are, of course, many nonobservable attributes of introversion and extraversion which can also be explained in terms of conditioning but, since it is not possible for external observers to rate such characteristics, they are of no direct relevance to the present discussion. If, therefore, only the externally observable characteristics of extraversion-introversion are considered, testable deductions can be made, deductions such as "Introverts form conditioned eyeblink responses more quickly than extraverts."

Problem

Testing such a prediction presents many difficulties of a methodological, technical and theoretical nature which have been discussed by the senior author elsewhere[*e.g.* 3, 4]. Even under idealized circumstances, such as the existence of a perfect measure of objective observable behavior, one might still not expect to find a perfect relationship between levels 2 and 3 as manifest in laboratory conditioning and observed behavior respectively. Thus, in certain circumstances, the good laboratory conditioner might learn to form *extraverted* behavior patterns if the need were great enough, whereas the poor laboratory conditioner would find much greater difficulty in learning introverted patterns of behavior and hence would be more likely to display the more general predicted relationship between conditioning and observable behavior. Thus, the environmental situation, or the conditioning the subject actually receives, may be at least as important as his conditionability. This is very apparent in the case of the delinquent who could be an antisocial delinquent because he is a good conditioner who has learned the wrong rules, or an asocial delinquent because he is a poor conditioner who cannot learn any rules, good or bad.

It is thus apparent that, even with the assumption that a perfect measure of observable behavior is available, the investigation of predicted relationships between laboratory conditioning and observable behavior presents numerous difficulties. If, however, the device used to measure this observable behavior is itself questionable, then the hazards involved in testing even such a simple proposition as "Introverts form conditioned eyeblink responses more quickly than extroverts" become enormously increased.

Both Eysenck and his colleagues, and those who wish to evaluate his theory independently are, for the most part, currently using the MMPI (earlier studies, including those of Eysenck and the present authors, have used the scales of Guilford or Cattell or parts of the MMPI or the Cornell Index, but essentially the same argument applies to these questionnaires if used to measure objective aspects of behavior). The MPI consists of two orthogonal self-rating scales, one measuring extraversion-introversion (E) and the other measuring neuroticism (N). The E scale may best be described as a measure of social tendency[5]. The items were devised by elaborate procedures involving factor and item analyses of other personality inventories, particularly the Guilford inventory of factors STDCR and the Maudsley Medical Questionnaire. The E and N scales have high "construct validity", that is, the items making up each scale are highly correlated with the factors they purport to measure and insignificantly correlated with the other factors.

Some of the items in this inventory (*e.g.*, "Are you inclined to keep in the background on social occasions?") could be rated by external observers, others (*e.g.*, "Are your daydreams frequently about things that can never come true?") can not be rated by external observers. For those self-rating items that can be fairly readily translated into a behavior rating form, a high agreement between the data obtained from the two forms might be expected. If Eysenck's theory of the causal relationship between laboratory conditioning and observed social behavior is pertinent then, as long as the methods used are equally objective and reliable, it should pertain irrespective of the method used to measure this observed social behavior. A completely objective method is presumably one in which, among other attributes, the subject is unaware of what is being measured and unable to influence the score in any way. In these respects, ratings by others are considerably more objective than self-ratings. Conversely, it might be argued that, for obvious reasons, the self-rater is sometimes in a better position to provide accurate information about himself than an external observer. Thus, quite apart from considerations such as the notorious unreliability of Yes/No questionnaires and complicating matters such as the presence of response set, acquiescence and social desirability factors, there are many reasons why the two kinds of ratings might not agree. Irrespective of the reasons why, if these two methods of measuring identical items of social behavior yield scores which differ appreciably, then serious difficulties arise in the testing of even the simplest deductions involving level 3.[2] Unless he is prepared to establish different predictions for each beforehand, the investigator is faced with the problem of which set of data to accept, those obtained from self-ratings or those obtained from ratings by others. It is suspected that some of the equivocation arising out of the attempts to investigate Eysenck's theory stems from such sources; in particular, the reliance on self-rating personality questionnaires as adequate measures of objective behavior.

Method and Procedure

The Questionnaire. A 40-item questionnaire was prepared in two forms, one suitable for self-rating and one suitable for rating by others. Each questionnaire, differing only in a translation from the second to the third person singular, consisted of 16 E items, 14 N items and 10 buffer items. These 30 E and N items were selected from the 48 items of the MPI solely because they were directly amenable to external rating procedures. Unlike the original MPI, in which a ? response is permitted, the present questionnaires were of the Yes/No variety, one point being awarded for each keyed response.

Subjects. One group consisted of 9 young, male alcoholics, all resident members of a 40 bed clinic who had been meeting regularly as a small unit for intensive group psychotherapy for at least three weeks. None of the Ss were illiterate, none were demonstrably psychotic or brain damaged. All were of adequate intelligence and educational background.

The other group consisted of 9 apparently healthy male students, all residents together in the same fraternity house for at least two years.

Procedure. Each group was tested separately, all S's being given first the self-rating form of the questionnaire. Upon completion with no time limit, each S was then asked to complete, one at a time and in counterbalanced order, one rating on each of his 8 comrades. No S knew the precise purpose of the investigation. Contamination was avoided by making the one testing session brief and closely proctored, and by removing each completed questionnaire before S was given the next one. Most S's completed all 9 questionnaires in about 2 hours. Candid responses were en-

[2]This lack of agreement would, of course, present no problem in those personality theories which utilize questionnaires to measure self-image or the so-called psychodynamic aspects of personality functioning, since both self and "other" ratings could be argued to be measuring different, but perhaps equally valid, aspects of the same person.

couraged by stressing anonymity and the research nature of the project. All S's were volunteers. For other purposes, not of concern here, a brief set of sociometric questions was appended to each questionnaire. In these questions S was asked to establish a preference rating, or selection from among his colleagues, for future shared experiences in a variety of appropriate situations.

Data Analysis. The E and N raw scores for each group were arranged in 9 × 9 score matrices with one row for each S judged and one column for each S acting as a judge. The self-ratings were transposed from their original diagonal cell locations to their respective row positions to form a tenth column which could then be treated as a

TABLE 1. CORRELATION MATRICES FOR E AND N FOR PATIENT AND NORMAL GROUPS

	SELF	A	B	C	D	E	F	G	H	I
					NORMAL GROUP					
SELF	—	.73	.41	.85	.77	.42	.90	.89	.86	.91
A	.29	—	.53	.83	.69	.53	.89	.60	.88	.86
B	.45	.43	—	.27	.51	.34	.43	.21	.26	.59
C	−.01	.32	.11	—	.65	.23	.86	.84	.84	.81
D	.32	.52	.14	.58	—	.68	.70	.66	.75	.78
E	−.32	.20	.52	.21	−.10	—	.43	.34	.61	.58
F	.43	.91	.56	.41	.58	.10	—	.71	.92	.92
G	.49	.50	.41	.12	.20	.42	.31	—	.79	.79
H	.31	.82	.51	.48	.62	.27	.77	.56	—	.87
I	.44	.20	−.04	.45	.12	−.03	.24	.50	.27	—
					PATIENT GROUP					
SELF	—	.14	.07	−.15	.12	−.20	.12	−.13	−.29	−.41
A	−.06	—	.62	.67	.87	.81	.85	.86	.67	.54
B	.11	.24	—	.56	.76	.83	.82	.66	.78	.28
C	.24	.01	.06	—	.56	.75	.73	.89	.79	.41
D	−.16	.34	.56	.18	—	.83	.85	.71	.63	.28
E	.39	.02	.51	−.14	.56	—	.94	.86	.93	.55
F	−.03	.48	.55	−.05	.90	.53	—	.84	.82	.48
G	−.40	.06	.17	−.14	.21	−.14	.26	—	.79	.68
H	−.29	.21	.34	.02	.60	.56	.46	.48	—	.47
I	−.16	.58	.21	.15	.71	.04	.82	.46	.29	—

Note: Above diagonals, E; below diagonals, N.

tenth set of judgments. The resulting empty cell in each column array was given the mean of all other judgments in the row identified by the empty cell and a Pearson product moment intercorrelation matrix obtained for each of the four sets of data (Table 1). Each matrix was then factor analyzed by the method of principal components and found to contain just one factor per matrix. The covariance loadings on the respective factors are presented in Table 2.

TABLE 2. COVARIANCE FACTOR LOADINGS

	Normal		Patient	
	E	N	E	N
SELF	3.15	1.22	−0.17	−0.13
A	2.55	0.82	4.64	1.44
B	1.50	1.28	3.88	1.80
C	2.34	0.55	3.78	0.14
D	1.68	0.89	3.24	3.00
E	1.46	0.26	4.02	1.25
F	2.64	1.26	5.08	3.32
G	1.93	0.85	4.14	1.09
H	3.08	1.16	3.82	2.04
I	2.78	0.51	2.17	2.52

Results and Discussion

As might be expected, each set of data yield one common factor suggesting that, whatever it is the raters are agreeing about, there is only one school of thought underlying this agreement. In the normal group the "self-rating" is indeed the best single estimate of both the E and the N factor (if this is, not unreasonably, assumed to be what the unitary factors designate). In the patient group the "self-rating" provides the worst estimates of the respective factors. Individually many of the other raters do better, collectively they provide a much better estimate of the factor. Whereas in the normal group the "other" judges tend to be better at judging E, in the patient group they tend to be better at judging N. The correlations between the ability to judge others on E and the ability to judge them on N are greater for the normal group than for the patient group, *i.e.*, the patient group as judges are less consistent from scale to scale than are the normal group.

From these observations, and in particular from a plot of the data presented in Table 2, it may be seen that, if the MPI scales (as modified here) are used to obtain estimates of a person's self-rating E and N scores, then the agreement of these self-rating scores with "other judges" scores is likely to be high for normal subjects but low in the case of certain patient populations. Thus, if the investigation of Eysenck's hypothesis is confined to carefully selected normal subjects using carefully selected "behavioral" items, then self-ratings are apparently as good as any careful judgment by others. However, when patient or neurotic groups are used (as in many instances) then a problem of questionnaire validity may arise since there may be no agreement between self-ratings and the judgments of others.

It might be argued that, since the patient group had known each other for only three weeks, any discrepancy between a self-rating and a rating by another could be attributed to a lack of knowledge by the other of the person being rated; therefore, one might expect the self-rating to provide the more accurate information. Such an argument is not tenable since, despite a group acquaintanceship of only 3 weeks, the "other" raters agree *among themselves* in their independent ratings of any one individual. (The readily testable assumption here is that if the above procedure were adopted with a group of strangers no similar common factor would be found.)

The present findings may explain the failure of Sigal, Star and Franks[9] to differentiate hysterics, dysthymics and normals in a satisfactory manner by means of the MPI used as a self-rating device. It seems likely that, despite its face validity, the MPI is not a good measuring instrument of objective behavior when used in the advocated self-rating manner and with mentally or emotionally abnormal S's. In the light of the good agreement among independent judges a very real problem of self-rating questionnaire validity and utility arises when such questionnaires are used to provide objective measures of behavioral characteristics. (It should be observed that this criticism of self-rating questionnaires necessarily need apply neither to questionnaires used clinically as projective devices to obtain self-image concepts, nor to self-rating questionnaires of the MMPI variety in which the criterion is simply their demonstrated ability to differentiate among various diagnostic groups.) It is possible that many of the equivocal or negative findings—and perhaps, even some of the positive results, obtained in investigating predictions relating Eysenck's levels 2 and 3 within disturbed populations may be in part attributable to the fact that all these workers used self-rating questionnaires to measure E and N. In such cases, perhaps the judgments of external raters, or even actual situational observations, might provide better measures of the E and N factors.

In conclusion, the clinical and research possibilities of utilizing a combination of self- and "other" rating techniques should be indicated. Further research is required to relate variables such as pathology, group homogeneity and length and type of acquaintanceships to observed discrepancies between the two kinds of ratings for various dimensions of personality. It is also possible that the differences in the two kinds of scores might lead to fruitful hypotheses about possible areas of disturbance or conflict. The differences in reliabilities and stability of the two scores

might also be of value in this respect. Another use might be in the area of psychopharmacology, since there is already some evidence[8] to suggest that certain drugs tend to increase both E type scores as measured by self-ratings and also the response set to reply "Yes" irrespective of the question being asked. Thus, such a drug might be expected to modify items scored in the "Yes" direction more when behavior is self-rated than when this same behavior is rated by others.

Summary

By a factor analytic method, it was found that self-rating measures of introversion-extraversion and of neuroticism (using a modified version of the Maudsley Personality Inventory) agree closely with measures obtained by external raters, the self-rating technique providing the best measures of the respective factors. In a psychiatric population no such agreement was found, the external raters agreeing among themselves but not with the estimates obtained by self-rating. From these observations it is suggested that there may be serious objections to the use of self-rating questionnaires for the measurement of observable personality attributes of behavior in abnormal populations. This possibility is of particular relevance to those concerned with the investigation of predictions derived from Eysenck's "stratification" theory of behavior.

References

1. Becker, W. C. The matching of behavior rating and questionnaire personality factors. *Psychol. Bull.*, 1960, *57*, 201-212.
2. Eysenck, H. J. Levels of personality, constitutional factors, and social influences: an experimental approach. *Int. J. soc. Psychiat.*, 1960, *6*, 12-24.
3. Franks, C. M. Some fundamental problems in conditioning. *Acta Psychol.*, 1958, *14*, 223-246.
4. Franks, C. M. Conditioning and abnormal behavior. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961, 457-487.
5. Jensen, A. R. The Maudsley Personality Inventory. *Acta Psychol.*, 1958, *14*, 314-325.
6. Laverty, S. G. and Franks, C. M. Sodium amytal and behaviour in neurotic subjects. *J. Neurol. Neurosurg. Psychiat.*, 1956, *19*, 137-143.
7. *The Manual of the Maudsley Personality Inventory*. London: Univ. London Press, 1959.
8. Scheier, I. H. and Cattell, R. B. Confirmation of objective test factors and assessment of their relation to questionnaire factors: A factor analysis of 113 rating, questionnaire and objective test measurements of personality. *J. ment. Sci.*, 1958, *104*, 608-624.
9. Sigal, J. J., Star, K. H. and Franks, C. M. Hysterics and dysthymics as criterion groups in the study of introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, *57*, 143-148.

A SAMPLE SURVEY OF MENTAL HOSPITAL PATIENTS[1]

E. I. BURDOCK, LEAH GLASS, ANNE S. HARDESTY, AND YVONNE M. BECK

Biometrics Research, N. Y. State Dept. of Mental Hygiene
and
Dept. of Psychiatry, Columbia University

Problem

A pilot study was conducted at New York State Psychiatric Institute to test the feasibility of sample survey methods for the evaluation of a mental hospital program. It involved an analysis of the hospital experience of patients as well as an examination of the disposition of hospital resources and personnel directed toward the recovery of patients. The focus of interest has been the individual patient— how his time is spent in various activities, how many and what kinds of treatments he receives, how much time he spends in the hospital, what privileges he has, how severe his illness is as reflected in his behavior on the ward, etc.

[1]This research was supported in part by a Public Health Service Research Grant (M-3546) from the National Institute of Mental Health.

Previous studies of psychiatric facilities have dealt with such questions as the preferability of a therapeutically oriented hospital setting as compared with a custodially oriented one[9], the social structure of the hospital itself or characteristic features significant for the effective functioning of the hospital as a treatment facility[3, 7, 11]. But psychiatric studies, for the most part, have not been concerned with the representativeness of their samples nor have they taken advantage of concomitant observations to determine the combined effects of competing influences. Statistical reports[10] provide an overall picture of patients by age, sex, diagnosis, race, admission status and disposition. However, questions about the effectiveness of a mental hospital program cannot be answered by examining routinely gathered statistical data because the relevant information is lacking. To collect the necessary information by taking a census of the total population of patients would be prohibitively expensive. Moreover, a complete enumeration would still be only a sample taken at one point in time and subject to error[6]. But some of these questions might be answered by surveying a representative sample of the population[8, 12].

The advantages of a sample survey over a complete enumeration or census of the population are: (a) Timeliness of the data because of the rapidity with which a survey can be completed. (b) Increased precision because of more careful attention to collection of data at the source instead of from existing records designed for other purposes. (c) Comprehensiveness because of the simultaneous collection of data about variables which have important effects upon one another. (d) Representativeness because collection of data need not span a prolonged time period with consequent confounding of seasonal fluctuations.

When the aim is to gain insight into the causes that govern the observed frequencies so that they can be controlled in time to come, small samples taken at frequent intervals may give much more information by revealing trends and changes as they occur, than either a total count or a large sample taken at wider intervals, and often for the same cost[5]. Thus, the U. S. Bureau of the Census has found it both economical and practical to obtain reliable estimates of socioeconomic variables by collecting supplementary data from a 25% sample instead of from the whole population. The construction of an adequate sampling design is therefore a major enterprise in its own right. Such a sampling design would, of course, include provision for estimating the sampling error.

The first task is to identify the critical sources of information in the hospital structure so as to determine the appropriate sampling units. Pilot studies are then necessary to obtain clues for reducing systematic error and for increasing the efficiency of sampling. Alternative sampling methods might be tried to ascertain whether simple random sampling, stratified sampling—with or without proportional allocation, or cluster sampling lead to the optimum combination of precision with economy. The report that follows presents only the frequencies of occurrence of individual characteristics. Analysis of their interactions is still incomplete.

Definition of the Population

A sample was selected from the population of resident patients in the hospital register. The population was defined as all resident patients of record at the beginning of an arbitrarily designated observation week. New admissions during the week were added to the roster, but releases were not dropped. In this way a total population of 155 patients was identified, all of whom were registered as residing at the hospital during all or part of the observation week. Information pertaining to specific behavior and activity on the ward and to treatment was gathered retrospectively for the observation week. Observations of patients and their records were made at the end of the observation week in order to avoid the possibility that either the patients or hospital personnel would so alter their behavior as to bias the study.

Selection of the Sample

The population was stratified according to sex; admission status (1st admission vs. readmission); and duration of current hospitalization (less than 6 months vs.

over 6 months). The population was found to be evenly divided between the two sexes; two-thirds of the patients were first admissions; half of the patients had been in the hospital less than six months. A twenty percent probability sample of 31 cases was then composed by selecting random subsamples from each of the eight categories in proportion to their size. This was the largest sample that could be studied because of the limitations of time and staff. An effort was made to assure a minimum of two subjects for each category. As it turned out, one patient was erroneously classified as a readmission and had to be shifted so that the category, readmitted males with over six months of hospitalization, ended up with only one individual. Table 1 shows the composition of the sample.

TABLE 1. CHARACTERISTICS OF POPULATION AND SAMPLE

	Population	Sample
First Admissions (Male)		
In hospital less than 6 months	17	4
In hospital over 6 months	29	7
Readmissions (Male)		
In hospital less than 6 months	20	3
In hospital over 6 months	8	1
First Admissions (Female)		
In hospital less than 6 months	20	4
In hospital over 6 months	24	5
Readmissions (Female)		
In hospital less than 6 months	23	5
In hospital over 6 months	14	2
Total	155	31

SOURCES OF INFORMATION

Data for the study were gathered from the case records, including nurses' notes, doctors' notes, and other clinical notations found in the records. Other written records such as doctors' order books, patients' sign-out books, lists of patients' appointments with doctors, etc., were also abstracted. In addition, nurses were asked to fill out the *Ward Behavior Rating Scale* [1] for each patient in the sample.

FINDINGS

Some of the proportions that follow are based on less than the total sample because the data were missing or the category was only applicable to part of the sample.

Admission Status. Almost all the patients in the sample were recorded as voluntary (in one case the relevant datum was missing). The majority of the patients (81%) had been admitted directly from the community; 13% were from receiving hospitals, while 6% had been transferred from other chronic treatment facilities.

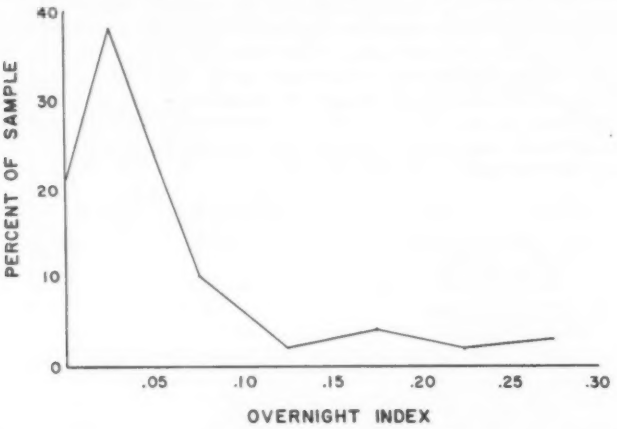
Disposition. During the observation week the discharge rate was 3%. Movement of patients from one ward to another was also at the rate of 3%.

Time out of the Hospital. Nineteen percent of the patients in the sample were either employed outside the hospital or attending school.[2] Thirty-five percent had spent two or more nights out of the hospital; 67% were allowed out on unaccompanied walks; 80% had the freedom of the building; 83% had special pass privileges either for outside work, for school, or for home visits.

An *Overnight Index* was computed for all patients in the study who had been in the hospital at least one week by taking the ratio of number of nights slept out to total number of nights of current hospitalization. Seven percent had been in hospital less than one week. For the rest the index ranged from zero to .26. Fig. 1 shows the frequency distribution for the Overnight Index. Twenty-one percent had an

[2] An additional 16% attended school on the hospital premises.

FIGURE 1. FREQUENCY POLYGON SHOWING PERCENT OF PATIENTS IN SAMPLE WHO SLEPT OUT OF THE HOSPITAL FOR VARIOUS PROPORTIONS OF TOTAL HOSPITAL STAY.



Overnight Index of zero, *i.e.*, they had slept in the hospital every night; 38% had slept out less than 1/20th of the time; 10% had slept out more than 1/20th but less than 1/10th of the time; 27% had slept out more than 1/10th but less than 1/4th of the time; 3% had slept outside for over 1/4th of the time.

Visitors. Thirty-seven percent of those for whom the data were reported had at least one visitor during the week; 30% had no visitors but spent at least one night out during the week. The remaining 33% neither spent a night at home nor had a visitor.

Security. Six percent of the patients were in a maximum security ward; 16% were in wards that were locked all the time; 13% were in the children's ward which was locked all the time; 65% were in wards that were open part or all of the time. During the observation week, 10% of the patients were under special observation because of suicidal tendencies or attempts to escape; 3% were under restraint.

Therapy. Ninety-seven percent received psychotherapy at least once during the observation week. Sixty-eight percent of the sample were recorded as attending ward meetings. Nineteen percent were reported as attending group psychotherapy sessions. Sixty-six percent participated in occupational therapy; 7% received hydrotherapy; and 52% attended gym. Half of the children attended other group therapeutic sessions, such as dance, art, and music therapy.

Drugs. Because of the great variety of drugs and the varying dosages, a single day was selected in the middle of the observation week on which all drugs and dosages for each member of the sample were recorded. According to the records 29% of the sample received no drugs at all; 29% were on one drug; 16% were on 2 drugs; 10% were on 3 drugs; 3% were on 4 drugs; 13% were on 5 drugs. For the 71% on drugs, the average number of different drugs administered during the day was 2.3. Altogether 30 different drugs were in use with the patients. Table 2 shows details of drug treatment for the observation day, the drugs being classified according to the pharmacological categories recommended for psychotherapeutic drugs by the AMA Council on Drugs.[3] Twenty-three percent of the sample received sleeping pills in the form of Doriden or chloral hydrate. Thirty-two percent of the sample were receiving phenothiazines: Thorazine, Compazine, or Stelazine, with Thorazine being administered to 19% while the newer analogues were administered to 13% of the patients.

[3]As reported in *JAMA*, 1958, *166*, 1040-1041.

TABLE 2. DRUGS ADMINISTERED TO A RANDOM SAMPLE OF PATIENTS DURING A SINGLE OBSERVATION DAY

Drug group	Specific Drug	Proportion of Patients	Average Dosage
Nonselective Depressants	Aspirin	10%	620 mg.
	Chloral hydrate	10%	517 mg.
	Doriden	13%	500 mg.
	Codeine sulphate	3%	15 mg.
	Dilantin	3%	300 mg.
Central Muscle Relaxants (sedatives)	Artane	10%	3 mg.
Central Sympathetic Suppressants (phenothiazine derivatives)	Thorazine	19%	408 mg.
	Compazine	3%	40 mg.
	Stelazine	10%	11 mg.
Central Nervous System Stimulants	Imprimine	13%	112.5 mg.
	Catron	3%	6 mg.
	Dexedrine	6%	7.5 mg.
Miscellaneous	Librium	6%	5 mg.
	SKF 556	6%	7.5 mg.
	WY 2445	3%	100 mg.
	Aspirin, Phenacetin and Ephedrine	3%	3 tablets
	Miscellaneous*	45%	....

*Includes neosynephrine, cortisone, colloidal gold, vitamins, colace, cogentine, feosol, gantrisin, Lugol's solution, etc.

Twenty-three percent of the sample were on stimulants: Imprimine, Catron or Dexedrine.

Physical Health. An examination of the medical records of the patients revealed that 13% had physical anomalies, 39% had at least one chronic physical illness, 10% had more than one chronic physical illness, and 10% had some acute physical illness. Three percent (8% of the women) were pregnant.

Background Variables. Background variables which other studies have shown to be related to the etiology or prognosis of mental illness were also tabulated. The ages of members of the sample show that it is a comparatively young population. Thirteen percent are children; almost half of the sample is under 25 years of age; almost half are between the ages of 25 and 44; only 6% are over 45. Fifty-nine percent of the adults are unmarried.

Table 3 shows that 83% are American-born, the remainder being either naturalized citizens or Puerto-Rican Americans. Sixty-three percent have at least one American-born parent; 13% are of Puerto Rican parentage, and 40% have one or both parents foreign-born.

Three percent of the sample were reported to be drug addicts. Six percent of the patients were alcoholics. Instances of mental disorder in the family were reported for 26%.

TABLE 3. PLACE OF BIRTH OF PATIENTS AND PARENTS

Nationality	patient	mother only	father only	both parents
American-born	25	3	2	14
Puerto Rican	2		1	3
Naturalized American	2			
Other	1	2	2	8
Not Reported	1			1
Totals	31	5	5	26

Socioeconomic considerations classify 21% upper class, 58% middle class, and 21% lower class. Twenty-seven percent were students. Forty percent of the women were classified as housewives.

Behavior leading to hospitalization. Table 4 shows the behaviors which led to the hospitalization of the patient. Depression was found in 42%, anxiety and fears in

TABLE 4. BEHAVIOR LEADING TO HOSPITALIZATION*

Behavior	Percent of Cases
Depressed state	42%
Anxiety, fears	35%
Confusion, disorientation	32%
Aggression towards self	29%
Aggression towards others	29%
Agitation and hyperactivity	26%
Delusions and Suspicions	29%
Withdrawal	26%
Affective display	29%
Physical complaints	26%
Habits (alcholism, drug addiction, etc.)	13%
Verbal peculiarities	13%
Hallucinations	13%
Bizarre behavior	10%
School behavior problems	10%

*The total exceeds 100% because more than one behavior was usually mentioned for a patient.

35%, and confusion and disorientation in 32%. Only 10% of the confusional states are described as organic. For children, the most frequently mentioned problems are those related to school behavior. This was true for three of the four children in the sample. Aggression towards self and others is also frequently mentioned.

Readmissions. Seven of the eleven readmissions in the sample were two years or more removed from the time of their first admission. For these seven an *Outcome Index*[4] was computed for their experience during these two years after first admission. The Outcome Index ranged from .02 to .93. Fifty-seven percent of these readmissions had Outcome Indices less than .125; 14% were between .125 and .50; the remaining 29% had Outcome Indices ranging from .50 to 1.0. Evidence from other studies[2] suggests that these figures may be interpreted as follows: the lower third are poor risks most of whom are likely to rotate in and out of hospital in future years; the upper third have a favorable prognosis for eventual adjustment in the community; prognosis for the middle third is guarded.

Severity of Illness. Scores on the *Ward Behavior Rating Scale*[1] ranged from 82 to 145.[5] Mean for the total sample was 120. Mean scores for males and females were about the same, as were those for patients with more or less than 6 months in the hospital. The largest difference was between first admissions with mean score of 122 and readmissions with mean score of 116, but it was not statistically significant.

The children in the sample were rated on a provisional form of the *Children's Behavior Rating Scale.* In order to correct for age differences, the scores were converted to proportion of responses showing good adaptation. The range was from 88% to 95% with a mean of 91%.

CONFIDENCE LIMITS

Because of the limitations of time and personnel and an error in the classification of the population which reduced one sampling category to a single subject, it was not possible to determine the sampling error empirically. Instead, theoretical

[4] $\frac{t_o}{l} \cdot \frac{r^2}{a^5}$ where t_o = time out in days; l = follow-up period since first admission in days; r = no. of releases; a = no. of admissions. The O. I. ranges from zero, never released, to a hypothetical one.

[5] Higher scores indicate better adaptation; maximum possible score is 150.

confidence limits [4] were computed for each of the proportions by applying the 95 percent binomial limits to the conditional distribution of p, given n, the total number of cases in the corresponding class, and n', the number for whom the relevant datum was reported. Table 5 lists the upper and lower confidence limits for a selected group of proportions which were based on the total sample of 31 patients.

TABLE 5. CONDITIONAL 95% BINOMIAL CONFIDENCE LIMITS FOR SELECTED PROPORTIONS REPORTED IN SURVEY

Factors	Proportion obtained	Lower limit	Upper limit
Treatment or disposition			
Discharged during week	.03	.01	.27
Employed outside or attending school	.19	.09	.35
Slept out two or more nights	.35	.22	.51
On open ward	.65	.48	.77
Under special observation	.10	.03	.23
Under restraint	.03	.01	.27
Received psychotherapy at least once	.97	.85	.99
Attended ward meetings	.68	.52	.80
Received group therapy	.19	.09	.35
Background characteristics			
Physical anomaly reported	.13	.04	.28
Had chronic physical illness	.39	.25	.54
Drug addiction	.03	.01	.27
Alcoholism	.06	.01	.19
Age groupings			
Children	.13	.04	.28
Under 25 years of age	.48	.33	.63
Between 25 and 45 years of age	.45	.31	.61
45 years of age and over	.06	.01	.19
Behavior leading to hospitalization			
Depression	.42	.28	.57
Anxiety and fears	.35	.22	.51
Confusion, disorientation	.32	.18	.49
Aggression against self	.29	.16	.46

SUMMARY

A pilot study was undertaken to demonstrate the feasibility of sample survey methods for the analysis of the application of hospital facilities to the care of mental patients. The method of selecting the sample and the sources of information are described. The report includes substantive findings regarding the characteristics of the sample and the population, together with estimates of the theoretical confidence limits.

REFERENCES

1. BURDOCK, E. I., HAKEREM, G., HARDESTY, Anne S., and ZUBIN, J. A ward behavior rating scale for mental hospital patients. *J. clin. Psychol.*, 1960, *16*, 246-247.
2. BURDOCK, E. I. and HARDESTY, Anne S. An outcome index for mental hospital patients. *J. abn. soc. Psychol.* In press.
3. CAUDILL, W. *The Psychiatric Hospital as a Small Society*. Cambridge, Mass. Harvard Univer. Press, 1958.
4. COCHRAN, W. G. *Sampling techniques*. New York: Wiley, 1953.
5. DEMING, W. E. *Some theory of sampling*. New York: Wiley, 1950.
6. DEMING, W. E. and STEPHAN, F. F. On the interpretation of censuses as samples. *J. Amer. stat. Assn.*, 1941, *36*, 45-49.
7. GREENBLATT, M., LEVINSON, D. J. and WILLIAMS, R. H. *The Patient and the Mental Hospital*. Glencoe, Ill.: Free Press, 1957.
8. HANSEN, M. H., HURWITZ, W. N., and MADOW, W. G. *Sample Survey methods and theory*. New York: Wiley, 1953. 2 vols.
9. JONES, M. *Social Psychiatry: A study of therapeutic communities*. London: Tavistock Publications, Ltd., 1952.
10. New York State Department of Mental Hygiene. *1960 Statistics of State Mental Hospitals*. Mimeo. Albany: Dept. of Mental Hygiene, Office of Statistical Services, 1960.
11. STANTON, A. H. and SCHWARTZ, M. S. *The Mental Hospital*. New York: Basic Books, 1954.
12. STEPHAN, F. F. and MCCARTHY, P. J. *Sampling Opinions*. New York: Wiley, 1958.

A BEHAVIORAL DISTURBANCE INDEX FOR PSYCHIATRIC PATIENT AND WARD DISTURBANCE

JAMES L. FRAMO AND ARTHUR M. ADLERSTEIN

Eastern Penna. Psychiatric Institute
Philadelphia, Pa.

Phila. Child Guidance Clinic
Philadelphia, Pa.

Problem

The Behavioral Disturbance Index (BDI) was developed as an attempt to categorize and quantify the pathological behavior of patients on a psychiatric ward, on both an individual and group basis. It is based on the systematic use of the routine nurse's notes which are recorded in practically every psychiatric institution. The BDI was originally created for a project done by Rashkis and Wallace[3] on the relationship between patient disturbance and staff attitudes. This is a report on the BDI, its description, range of applicability, reliability, validity, shortcomings and possible usefulness to those engaged in clinical research.

Lorr's[1] excellent review article on "Rating Scales and Check Lists for the Evaluation of Psychopathology" distinguishes between scales for use by psychiatric aides and nurses and scales for use by trained clinicians. Standard psychiatric rating scales based on interviews are very serviceable for quantifying clinical judgments but they frequently yield scores based on guarded "office behavior" or staff conference behavior which may be quite at variance with the patient's everyday functioning on the ward. Scales completed by nurses and aides introduce other problems. Ward personnel may become oversensitized to certain aspects of behavior while ignoring others. Observations may be contaminated by such prestige questions as to who will be the evaluator of the ratings. There is no way of estimating in what way ward personnel affect the very behaviors being observed and evaluated.

The BDI attempts to obviate these limitations and to yield a naturalistic measure of patient disturbance which more adequately measures functioning under living conditions of the ward. The spontaneous observations which are made most commonly and routinely are the daily nurse's notes. Although psychiatric nurses and aides are generally required to make routine observations of patient behavior three times a day, these notes have rarely been used in any systematic way. It is recognized that there is considerable variability in quality and inclusiveness of nurse's notes from hospital to hospital, but, assuming some minimal standards of reporting, they contain a valuable source of meaningful information.

Method and Rationale

The method used in determining the items of the BDI was inductive and empirical. The earliest stage of development consisted of reading more than 4000 nurse's notes and from these were culled a wide range of symptoms, behaviors, and incidents. The categories underwent continuous modification as they were applied to subsequent patient evaluations. The BDI went through six revisions before the current list of 25 items was selected. These 25 items include practically all of the behaviors likely to be reported in nurse's notes.

Most of the items include two sub-categories which describe disturbances at both ends of an item bi-polar construct. It was our belief that extremes of any kind should be viewed as disturbed behavior even when such behavior is not ordinarily considered pathological, or, is, in fact, generally socially rewarded. Such items as overforbearance, liking group activity too much, indifference to bodily illness, prudishness, fastidiousness and over-cooperation in ward routines are of this nature. With some of the categories (*e.g.*, delusions) it was not possible to conceptualize an over-and under-reaction. The BDI differs from the usual psychiatric rating scale in that the emphasis is on psychopathological behavior; it does not allow for estimations of the "good" patient or of adaptive behavior. Psychiatric health, from this viewpoint, is considered simply as freedom from disturbance of the type which comes to

the attention of ward personnel. The validity study to be described will demonstrate that this does have real meaning. The 25 items of the BDI are given in Fig. 1.

FIG. 1. THE BEHAVIORAL DISTURBANCE INDEX

1. Aggressive disturbance (motoric)
   a. Against others
   b. Against self
   c. Against material objects
2. Aggressive disturbances (verbal)
   a. Against others
   b. Against self
   c. Against material objects
3. Bizarre behavior
4. Complaining behavior
   a. Whining, complaining
   b. Overforbearing; self-denial
5. Compulsive disturbances
6. Delusional disturbances
7. Eating disturbances
   a. Refusal to eat
   b. Overeating
8. Feeling disturbances
   a. Panic, anxiety attack, extreme fear
   b. Flat, apathetic, dull
   c. Inappropriate affect
   d. "Nervous," anxious, upset
9. Hallucinatory disturbances
10. Manners disturbances
    a. Unconcerned with social amenities
    b. Overpolite, obsequious, ceremonious
11. Mood disturbances
    a. Elated, exhilarated
    b. Depressed
12. Motor disturbances
    a. Overactive, pacing, running, agitated
    b. Immobile, catatonic, posturing
13. Paranoid disturbances
14. Participation disturbances
    a. Refuses to go to OT or group activities
    b. Likes group activity too much; too active
15. Privilege disturbances
    a. Irresponsible privilege behavior
    b. Refuses privileges
16. Physical Complaint disturbances
    a. Complains of headaches, stomach aches, etc.
    b. Indifferent to bodily injury, illness
17. Responsivity disturbances
    a. Overdependent, clinging, helpless
    b. Negativistic, resistant, stubborn, sullen
18. Sex disturbances (both homo- and heterosexual)
    a. Masturbation, exhibitionism, seductiveness
    b. Prudish, disapproving of sexual activities
19. Sleep disturbances
    a. Oversleeping
    b. Insomnia; has trouble sleeping
20. Sociability disturbances
    a. Seclusive, withdrawn
    b. Forces self on people; a "pest"
21. Thinking disturbances
    a. Confused, disoriented
    b. Perplexed, preoccupied
22. Tidiness disturbances
    a. Sloppy in dress; untidy in toilet habits
    b. Overly fastidious in dress or toilet
23. Verbal disturbances
    a. Tangential, irrelevant speech
    b. Rarely talks, mute
    c. Loud, boisterous, shouting, overtalkative
24. Visitor disturbances
    a. Overanxious about visitors
    b. Negative about visitors
25. Ward routine disturbances
    a. Does not cooperate in a ward routine
    b. Overcooperates, seeks work unnecessarily
26. Miscellaneous (write in)

SCORING

A manual has been developed for the BDI which includes general principles of scoring, and descriptions and examples of each of the items.[1] The general principles of scoring from the manual are as follows:

1. Inasmuch as the focus of interest is on disturbance, appropriate, healthy, trivial or neutral occurrences and behaviors are not checked on the BDI. In general, if a behavior or symptom is striking enough to be remembered and noted by the nurse or aide in her clinical notes, that disturbance should be checked if it belongs under one of the BDI categories. (The following are nurse's notes which would not be checked, "Patient seems to be happy today." "Patient outgoing and conversed with staff today." "Patient received a package today." "Patient had a visitor today.")

2. In complex behaviors, the separate items are checked, e.g., if nurse's note says "Patient complained of stomach-ache, refused to get out of bed and ate little at breakfast", all of these items are checked.

[1]Considerations of space prohibit reproduction of the entire manual. It may be obtained from the senior author.

3. If the same behavior occurs in successive shifts, it is checked provided the nurse or aide is not reporting hearsay information (*i.e.*, a nurse reports that she heard from the nurse on the first shift that the patient had a fight.)

4. Only disturbances which are reported are checked; no inferences should be made. If patient stayed in bed all day, just that item is checked on each shift and not other disturbances such as participation disturbances, etc., which are a concomitant of this behavior.

5. If the same behavior occurs in the same shift as part of an unbroken pattern it is considered a single disturbance. Whereas, if its continuity is broken, i.e., an event or incident intervenes, it is checked as frequently as this occurs. (E.g., patient complaining of a headache continuously receives one check in this category, but if patient complains of a headache, gets into a fight, and complains of a headache again he gets two checks for Physical Complaint even though these occur during the same shift.)

6. Nurses should be discouraged from writing such things as "No change in behavior," "Patient's usual behavior." When this does occur, as it sometimes does, the observer should check back to the past previous notation which stated what that usual behavior was and use his judgment as to what categories to check.

The actual scoring of the Index is simple and straightforward: the recorder reads through the nurse's notes for a particular period, say, one day, and makes a check opposite the appropriate item on the list. Checks are entered for each of the shifts on which the behavior or incident is noted. Behavior for which there is no category on the BDI may be described under Miscellaneous; it is our experience that this very rarely has to be used. In the initial application of the BDI we were interested in the amount of disturbance manifested by patients over a one week period. A form was developed which provided for entries opposite each disturbance. This allowed for daily entries and a weekly total. A space was provided at the bottom of the form for the total number of disturbances observed for the week. This procedure can, of course, be modified for the purposes of a particular research problem.

Reliability

Inter-rater reliability was obtained by having two raters independently score the disturbances of 20 patients for a one week period. Reliability was derived by the use of two correlation techniques. The first method was to correlate the two sets of "total" scores; this was done to determine how much agreement there was between raters as to the gross amount of disturbance. The correlation was .97, significant well beyond the .01 level. This suggests that two raters trained in using the BDI will pick up almost all of the disturbances noted by ward personnel.

This correlation, however, says nothing about agreement on *particular* items; it is theoretically possible to have agreement on total disturbance when in fact the raters have used entirely different categories. We therefore determined inter-rater reliability for the separate items for the week as well as the agreement on separate items for a particular day. The data presented a unique problem since *not* checking an item also constituted an agreement. In order to correct for this and to rule out a spuriously high correlation based on chance agreement, a formula was derived which took these factors into account.[2] The mean item inter-rater reliability for the week was .69 and the mean agreement for a particular day was .62 (an r of .428 is reliable beyond the .01 level). We can therefore conclude that inter-rater reliability is well within acceptable limits.

---

[2]We wish to thank Drs. B. Karon and D. Estavan for their help in deriving this formula. Given:

C_1 = The number of blocks rater one checks
C_2 = The number of blocks rater two checks
N = The number of blocks the test is composed of
P = The number of agreements (N − the number of disagreements)
P' = The number of chance agreements

$$P' = \frac{C_1 \quad C_2}{N} + \frac{(N - C_1)\,(N - C_2)}{N}$$

$$\text{Inter-rater reliability} = \frac{P - P'}{N - P'}$$

Validity

The validity of the BDI was assessed by comparing the scores of a group of discharged patients with the scores of a group of still-hospitalized psychotic patients. We hypothesized that while both groups would be equally disturbed at the beginning of hospitalization (*i.e.*, relatively equal on scores), the group of discharged patients should manifest less disturbance just prior to discharge than a comparable group of hospitalized patients.

Twenty female patients from the Eastern Pennsylvania Psychiatric Institute constituted the validation study group. Of the 20 patients, 10 had been discharged from the Institute at the time of the study and 10 were still in the hospital. Criteria for both groups included hospitalization for a period of at least one month, no history of shock therapy, and the absence of a diagnosis of central nervous system damage. In addition, the patients in the discharged group had to have been discharged by the staff (not voluntarily signed out) and had to have remained out of the hospital for at least six months after discharge. The hospitalized group had to have been in-patients for a period of at least three months after the last recording of the BDI scores. From a large pool of discharged patients, all of whom met the criteria, we randomly selected the study group of ten. Ten other patients matched on the relevant variables were then selected from the still-hospitalized group. Entries on the BDI were made for each patient from nurse's notes on the third and fourth day of hospitalization in order to allow several days for the patients to become acclimated to the hospital. Entries were also made for the discharged group from notes made three and four days prior to discharge. For the hospitalized group, entries were made using notes covering a comparable length of time. The mean length of the observation was 55 days for the discharged patients and 56 days for the hospitalized group. The two validation groups of improved and unimproved patients, then, were balanced on the parameters of sex, age, diagnosis, and length of hospitalization.

In order to test the hypothesis it was important to know if the two groups differed initially on BDI scores. The mean score for the still-hospitalized group was 26 and for the discharged group the mean score was 19; this initial difference in mean scores was not statistically significant ($t = 1.45$, $> .05$). In other words, patients in both groups were about equally disturbed upon entering the hospital. We then tested to see if there was a difference in scores for the second or final set of entries. The mean of the discharged group had dropped from 19 to 4.5 while the mean of the unimproved patients had dropped from 26 to 22 ($t = 3.70$, $< .01$). This clearly indicates that the BDI reflects judgments of the staff about readiness for discharge.

As an additional test we calculated the differences between each group's initial and final scores to see if the drop in mean scores represented a statistically significant drop. This was done by testing the difference scores in each group against a null hypothesis of zero drop. The difference t in the still-hospitalized group was 1.05, which is not sufficient to warrant rejection of the null hypothesis. The difference t in the discharged group was 3.70 ($<.01$). This indicates that the difference between initial and final scores for the unimproved group can be accounted for on the basis of a chance drop, indicating that there is no reason to believe that there was improvement in the still-hospitalized group. The difference for the discharged group probably represents a true drop in score or disturbance. This validation study, then, clearly substantiates the hypothesis that the Behavioral Disturbance Index is a valid measure of change in the ward behavior of patients.

Discussion

The Behavioral Disturbance Index has been demonstrated to be a reliable instrument which can differentiate between recovered and non-recovered psychotic patients. The validity study demonstrates that the readiness for discharge as decided by patients' psychiatrists is reflected in the lowered disturbance measured by the BDI.

In its present form the BDI points up only gross behavioral disturbance. It does not attempt to characterize patient disturbance psychodynamically and it is not possible to state with certainty that a higher score means greater psychopathology. This is not necessarily a disadvantage since it is well known that patients with severe inner pathology frequently manage to survive socially; the BDI appears to reflect the kind of overt disturbed behavior which comes to the attention of others socially. Society usually does not care how emotionally sick a person is unless they show it.

It is recognized that the total disturbance on the BDI is a summation of disparate behaviors which are certainly not equivalent. The problem of weighting presents many difficulties. It might be possible to weight the different items in accordance with some theory of psychopathology, or empirically. The approach of Phillips and Rabinovitch [2] who statistically interrelated patterns of symptoms is an example of empirical treatment of these kind of data. Their categories of "Avoidance of Others", "Self-indulgence and Turning Against Others", and "Self-deprivation and Turning Against the Self", appear promising as methods of ordering symptoms and behaviors into meaningful clusters.

Nurse's notes have the obvious advantage of availability; the research investigator can rely on existing data and go back to any point in time. For example, he can select a group that was given a particular form of treatment and compare pretreatment behavior with post-treatment behavior. He can also trace the movement of patients whose initial behaviors were similar and who experienced different courses of treatment. It might be possible to delineate from symptom clusters which patients are likely to respond to treatment.

The BDI was originally devised to evaluate patient disturbance on an experimental ward. Just as an individual patient's disturbance can be totalled, so can the disturbances of an entire ward be measured so as to reflect experimental or therapeutic changes introduced on a ward, or the effect of ward personnel shifts. Those workers utilizing therapeutic community techniques can get a quantified estimate of ward excitement or "tone". It should be remembered, however, that the BDI is more intended to enrich rather than replace other observations.

Summary

A Behavioral Disturbance Index, based on categorization of behaviors and symptoms obtained from nurse's notes, has been found to be a useful instrument for reflecting the degree of disturbance of an individual psychotic patient or whole wards of patients. This empirically derived method of evaluating patient or ward disturbance is relatively simple to use, proved to be reliable, and was valid as a reflection of the progress or chronicity of the hospitalized psychotic patient in the living situation of the ward. The BDI yields quantitative scores which can be related to any other relevant research data, such as the effect of treatment on an individual patient, or the effect on the total ward of introducing experimental or therapeutic regimes.

References

1. Lorr, M. Rating scales and check lists for the evaluation of psychopathology. *Psychol. Bull.*, 1954, *51*, 119-127.
2. Phillips, L. and Rabinovitch, M. S. Social role and patterns of symptomatic behaviors. *J. abn. soc. Psychol.*, 1958, *57*, 181-186.
3. Rashkis, H. S. and Wallace, A. F. C. The reciprocal effect: how patient disturbance affects and is affected by staff attitudes. *A. M. A. Arch. gen. Psychiat.*, 1959, *1*, 489-498.

CONSULTATION VS THERAPY IN THE PSYCHOLOGICAL TREATMENT OF NP HOSPITAL PATIENTS

MILTON B. JENSEN

Veterans Administration Center
Hot Springs, South Dakota

Introduction

In a preliminary study [1] we found that both group psychotherapy and consultation in behalf of male neuropsychiatric hospital patients were related to transfer from closed to open wards and to departure from the hospital via trial visit and medical discharge. Fisher's Exact Test p was .06 for 14 in the intervention groups (7 who received 25 hours of group psychotherapy and 7 in behalf of whom there were 21 hours of consultation) as opposed to seven for whom there was no psychological intervention. The data were secured 142 days after onset of the experiment and 74 days after its termination. There was no psychological intervention in behalf of any of the patients during the follow-up period. The data indicated, however, that our matching of patients in terms of prognosis for recovery was not readily defensible and might have prejudiced our findings in favor of the treatment groups. The present study represents a replication of our pilot study with more strict controls.

The experimental design for the present study was a joint production of many individuals—psychology staff, trainees and consultants in psychology from the University of Tennessee. The three assistants[1] conducted the consulting and psychotherapy as a part of their training at the Veterans Administration Hospital, Salisbury, North Carolina. At the time of the experiment, they were advanced trainees about to return to the University of Tennessee for their last year of graduate study. All had some training and experience in group psychotherapy, but this was their introductory training as consultants in a NP hospital setting. The writer met with them at least weekly throughout the experimental period to formulate plans and discuss problems.

The procedure aimed at modification of those behaviors considered responsible for closed-ward status and NP hospitalization. We hypothesized that: (1) Group psychotherapy with and consultation in behalf of NP hospital patients would lead to open-ward status and non-hospitalization, and (2) alternatively, consultation in behalf of patients would be followed by more such changes than would follow group psychotherapy. In formulating the second hypothesis it was reasoned that consultation relative to the patient with many of the hospital personnel would extend the influence of the psychologist and be effective by reason of something of a "total push." That is, the therapist spent only a few hours per week with the patients in group psychotherapy while nurses, nursing assistants, social workers and technicians spent many hours per day with them. It was thought that advising nurses and other personnel relative to altering the behaviors of patients would result in changes that would be reflected in open- rather than closed-ward status and in leaving and staying out of the hospital.

Method

Initially, 44 closed-ward, non-geriatric female NP hospital patients were randomized into three groups of 16, 15 and 13. At random one group was chosen for consultation (in behalf of), one to receive group psychotherapy, and the third as a non-intervention or control group. Other patients admitted to the same closed wards during the first 10 weeks of the 13-week experimental period were randomly assigned to one of the three groups and those who left the hospital during the first

[1]Assisted by William F. Gwynn, Ft. Hays State College; Zachery Mayberger and Garret H. Yanage, University of Tennessee. I am deeply indebted also to Dr. William E. Morris for assistance in devising this study and in preparation of this report.

two weeks of the study were excluded. The experimental period extended from May 5, 1959 to July 31, 1959, inclusive (88 days); the follow-up period from August 1, 1959 to December 31, 1959, inclusive (153 days).

All of the patient-subjects were members of one of the three groups throughout at least two consecutive weeks of the experimental period and throughout all of the follow-up period. The mean number of patient days for the three groups during the experimental period did not differ significantly. They were: I. Consultation group, 74.6 (SD, 24.3); II. therapy group, 68.2 (SD, 22.1); III. non-intervention control group, 77.0 (S.D., 17.6). The numbers of patients who spent at least two weeks of the experimental period in one of the groups varied because of chance factors and attrition through leaving the hospital.

The same physician was in charge of all the patients throughout the study and follow-up periods and the make-up of the ancillary personnel was relatively constant. Personnel other than psychology staff had no knowledge of the nature of the study nor even that a study was in progress. This secrecy was maintained while the patients were being selected and during the experimental and follow-up periods.

The psychotherapy group (II) met twice weekly for 13 consecutive weeks in sessions of about 90 minutes each from May 5, 1959 to July 31, 1959, inclusive. During this period the therapists spent about the same amount of time with nurses, nursing assistants, social workers and others (physical medicine and recreation technicians, etc.) in behalf of the consultation group (I). They gave no therapy to the consultation group during the experimental and follow-up periods and had no contact with the responsible physician relative to any of the patients in the study during that period. There was no psychological intervention in behalf of any of the patients in the study during the follow-up period.

The therapy of Group II proceeded along more-or-less conventional lines but with more-than-usual attention to specific behaviors. For example, one patient incessantly and in a loud voice berated the personnel because she was "locked up." Therapist and group efforts were directed towards getting her to use something of a "sales" approach in her efforts to get attention and to get transferred to an open ward.

Consultation in behalf of Group I was complicated by the fact that it was not sought by the consultees (nurses, *et al*) and that consultees had to be kept unaware of the criteria— open-ward status and de-hospitalization. We would have preferred a more normal consultant-consultee relationship but found no way of attaining it without violation of the secrecy we considered essential to experimental control. We deduce that secrecy was maintained from the post-study expressions of surprise of some of the consultees when informed that the study had been made.

As the study progressed it appeared that the therapists were having difficulty with the consulting role. They couldn't get in quite as much time consulting as they devoted to psychotherapy. They didn't always know how to structure the consultation—how to draw from nurses and others descriptions of critical behaviors, etc. They seemed to have some difficulty (mostly in the early stages) in centering consultation on behavior change and not on change in hospital status. We intended that treatment should center on behavior rather than on the criterion of changed status. We would have preferred that the therapists had no knowledge of the criteria, but we saw no way of accomplishing this. However, their knowledge of results was limited since the data were not analyzed until after completion of the study.

There was no known psychological intervention in behalf of Group III during the study. They were not taken into psychological therapy. No persons were approached or consulted in their behalf by psychology personnel. When the therapist was approached by hospital personnel concerning a member of the control group, he resorted to evasion and subterfuge to avoid influencing the questioner. Discussion of members of the control group by the therapist was scrupulously avoided in all therapy and consultation contacts and was circumvented if possible when introduced by others.

RESULTS

TABLE 1. LOCATION OF MEMBERS OF THREE GROUPS*

Groups	Closed-Ward	Open-Ward	Out of Hospital
Start of Experiment:			
I. Consultation	23		
II. Group therapy	32		
III. Control	20		
End of 88-Day Experiment:			
I. Consultation	12	6	5
II. Group therapy	12	11	9
III. Control	13	7	0
End of 153-Day Follow-Up:			
I. Consultation	6	5	12
II. Group therapy	5	10	17
III. Control	11	3	6

*Corrected Chi Square p (one tail) favors the treatment groups combined versus the control group in terms of being on open wards or out of the hospital as follows: .05 at the end of the experimental period and .004 at end of the follow-up period. Seemingly, some of the effects of psychological treatment become apparent a considerable period of time after cessation of the treatment.

The data support hypothesis 1 that psychological treatment via consultation (in behalf of) and group psychotherapy would lead to open-ward status and departure from the hospital. A trend in this direction at the .05 level of confidence was evident at the end of the 88-day experimental period (Table 1) and beyond the .01 level of confidence at the end of the subsequent 153-day follow-up period (Tables 1, 2 and 3).

TABLE 2. MEAN PER CENT OF TIME GROUPS SPENT IN VARIOUS CATEGORIES WITH STANDARD DEVIATIONS OF MEANS IN PARENTHESES

	N	A Closed-Ward	B Open-Ward	C Out of Hospital	B + C
I. Consultation	23	50.4 (7.8)	20.6 (5.9)	29.0 (6.8)	49.6 (7.8)
II. Group therapy	32	33.4 (4.2)	32.6 (5.9)	34.0 (6.2)	66.5 (5.7)
III. Control	20	69.1 (7.8)	13.6 (3.8)	17.3 (6.0)	31.0 (6.8)
I + II	55	40.6 (4.2)	27.6 (4.2)	31.9 (4.7)	59.4 (4.8)

The data do not support hypothesis 2 that consultation (in behalf of) would be more effective than group psychotherapy in moving closed-ward NP hospital patients to open wards and out of the hospital. In fact, there is a trend (not significant) in the opposite direction (Tables 1, 2 and 3).

TABLE 3. CRITICAL RATIOS OF MEAN DIFFERENCES FROM TABLE 2.

	A Closed-Ward	B Open-Ward	C Out of Hospital	B + C
I – II	1.92*	−1.45	−.33	−1.76
I – III	−1.69	1.01	1.29	1.49
II – III	−4.04**	2.69**	1.94*	2.88**
(I + II) – III	−3.23**	2.45**	1.92*	3.43**

*Considered significant or nearly so.
**Considered very significant

Conclusions

The data show that both consultation and group psychotherapy were related to change to open-ward and non-hospital status, but that group therapy was the more decisive factor. We think these changes resulted from changed behaviors and that the changes in behavior stemmed primarily from psychological treatment. However, our data do not properly define those behaviors. Unfortunately, we did not secure systematic data which would relate specific behaviors to being in the hospital and to being on closed as opposed to open wards. We saw too late that we had opportunity for this. An occasional note by the therapist says, for example, "Patient has not written his folks that the F. B. I. is after him in the past three weeks." We hope that others will investigate this area. Actually, we do not know which behaviors, if any, were responsible for NP hospitalization and/or ward status of any of our subjects. Perhaps a hierarchy of behaviors is responsible.

References

1. Gwynn, W. F. and Jensen, M. B. Direct, indirect, and no psychological treatment of NP hospital patients (A pilot study). Veterans Administration Hospital, Salisbury, North Carolina, 1959 (mimeographed). Presented at the meeting of the North Carolina Psychological Association, May, 1959.
2. Jensen, M. B. and Morris, W. E. Reliability-unreliability of ancillary psychiatric evaluations. *J. clin. Psychol.*, 1960, *16*, 248-252.

EMPIRICAL SYNDROME GROUPINGS OF SCHIZOPHRENIC HOSPITAL ADMISSIONS[1]

WILSON H. GUERTIN

University of Florida

Problem

The search for test correlates with schizophrenia is like the search for the Holy Grail; the object of the search may actually exist, but the only certain reality is in the minds of men. Conflicting proposals of relations between schizophrenia and behavior or test variables, and the failure to substantiate them upon cross-validation, usually are attributed to sample differences (8, p. 272-3). Since schizophrenia is generally acknowledged to be a collection of variously described syndromes (8, p. 277), there is little reason to expect uniform sampling of subjects for research, and even when it does occur, the obtained results would have little application for other different samples of schizophrenia. What clearly is needed to provide a break-through in research on schizophrenics is some form of quantitatively specified typology.

Armchair speculation about types of schizophrenia has persisted since Kraepelin's first proposals, with little general agreement among psychiatrists, and produced almost no worthwhile research findings. As we move from the armchair to a more quantitative study of similarities among individuals we first encounter linkage analysis (7). This approach, like the armchair speculation, fails to provide a unique solution to the problem of what are the particular types, *i.e.*, many alternative solutions are possible. Just as the incorporation of the simple structure concept into factor analysis provides a unique and finite solution in analyzing clusters of traits, unique solutions as to types of people in the sample can be derived similarly from clusters of people (linkage-types). Transposed factor analysis does this by finding the simple structure solution for a matrix of intercorrelations between people, expressing degrees of similarity with respect to such variables as psychiatric symptoms.

In their efforts to identify major psychiatric syndromes, factor analysts have seldom focused on the schizophrenias exclusively. As in optics, the smaller the field

[1]The author is greatly indebted to Drs. Frank du Mas, Richard Jenkins, Leonard Goodstein, and Sanford Golin for their many helpful suggestions. From the V. A. Hospital, Knoxville, Iowa.

area, the more apparent the detail becomes. Only two studies[2, 3] have employed behavior symptoms of schizophrenics in a transposed factor analytic design to get a glimpse at the underlying types of schizophrenics. Both these studies failed to describe the obtained type-factors statistically in terms of symptoms. This is a consequence of the transposed factor analytic methodology that gives results in terms of factor loadings for individuals rather than symptoms. Types derived in this fashion are, therefore, sometimes described verbally by referring to the symptoms held in common by the few heaviest loaded individuals on each factor. Or, more quantitatively, the relation of a symptom to a type-factor can be represented by a correlation or chi-square to express the presence of the symptom associated with individuals lying in the upper range of factors scores.

Neither of these methods is consistently satisfactory for demonstrating the relationship between presence or absence of a symptom and the individual's type-factor score. Pearson correlations will be low when only the most typical or least typical individuals show marked presence or absence of the symptom and there is no covariation throughout the remainder of factor-score range. Should the moderate factor loadings be associated with the presence or absence of a symptom, a chi-square based upon a single high cutting line will completely miss the relationship. Were the cutting line moved around to fit the obtained distribution for each symptom, a worthwhile description of the type-factor in terms of symptom occurrences associated with different ranges or levels of type-factor loading would result. However, there is no longer any need to proceed in such a haphazard manner since du Mas[1] introduced the systematized procedure for identifying relationships between symptom categories and such scale values as type-factor loading called Manifest Structure Analysis (MSA).

Thus, the present study seeks to illustrate the bringing together of factor analytic typing and MSA description of the types. Type-factors are to be identified from a sample of the symptoms of chronic male schizophrenics, and MSA *catescales* of symptoms as they relate to each of the obtained type-factors are to be established to make it possible to calculate type-factor loadings directly from symptom data through the use of tables. Or, more grossly, the methodology to be illustrated is directed toward providing quantified reference information that can serve as diagnostic criteria to provide homogeneity and reproducibility of *Ss* for studies relating variables to "schizophrenia".

Procedure

The 49 patients of this study were male veterans with hospital diagnoses of schizophrenia. The upper age limit was 40. Almost none of the cases were first admissions. Most were readmissions to the VA Hospital, Knoxville, Iowa, others had been previously hospitalized in other VA or state hospitals and were direct transfers or had failed in their efforts to readjust to the community. The clearly chronic nature of the sample would make it impossible to generalize data from this sample to an acute sample. Lobotomized patients and those suffering from post-EST confusion were excluded. Many were on ataractic drugs at the time of evaluation. The fiftieth subject was an hypothetical normal based upon the writer's estimates of item scores for a normal individual in a hospital situation.

The aim of the study, to identify types of schizophrenics, made it desirable to evaluate only those symptoms which were relevant to this differentiation. None of the most applicable rating scales[4, 6, 9] were designed to study schizophrenics exclusively, so a combination of items from all three was used to provide as many relevant items as possible. Nineteen of the originally selected items proved to have undesirable semantic features or their frequency of occurrence was higher than 90 or less than 10 per cent.[2]

---

[2]Copies of the orthogonal and oblique factor matrices, as well as the rating scale and tables for converting mean score values of individuals to factor loadings, have been deposited with the American Documentation Institute. Order Document No. 6559, remitting $1.25 for microfilm or $1.25 for photocopies.

All ratings were made by the author on a five-point scale during the first week after admission. Most of the information was obtained in an interview that lasted from 20 minutes to an hour. Additional information about ward behavior was obtained by interviewing a nursing assistant (attendant). Cutting lines were established and data transferred to keysort cards for calculating tetrachoric correlations between each of the subjects. The 50 × 50 intercorrelation matrix was factored by the multiple-group centroid method. The obtained factors were rotated to orthogonality and then to an oblique simple structure.

The principles of manifest structure analysis[1] were employed to demonstrate the empirical relationship between a given type-factor loading and the presence or absence of each of the symptoms. A catescale was established for each type-factor, taken one at a time, as follows:

1. *Ss* were ranked from low to high by factor loadings, and became the ordinate on the scaling frame.
2. Symptoms dichotomized into present or absent became the abscissa on the scaling frame in their chance order of arrangement in the rating scale.
3. Pins were inserted for a person in each cell of his designated row to correspond to the presence of that symptom represented by each particular column.
4. Non-scalable symptoms were rejected by removing their column slats from the frame. Some were retained as reciprocal categories[1].
5. The remaining column slats were placed in ascending order by mean rank on the ordinate scale in order to approximate the intensive mathematical model (identity matrix). In this state, the scaling frame constitutes an associational surface expressing the relationship between categories and type-factor loading.
6. A tabulation of the categories and their mean ranks now constitute the derived catescale.

Results

Estimated communality was 70.7 per cent of the total variance, thereby indicating that most of the symptom variance was identifiable in terms of the ob-

Table 1. MSA Scale Relating Symptom-Categories to Type-Factor: Normal (Coefficient of Reproducibility = .86)*

Scale Value	Brief Description of Symptom and Item Number
315	Very aware of other patients (77)
313	Very cooperative (68)
310	No obsessive thoughts (21 −)
303	Usually relevant (1)
303	Not opinionated (36)
280	Demands excessive attention (67)
239	Can verbalize hostility toward patients (76)
234	Sloppy appearance (63 −)
223	Seldom looks at E. (57 −)
201	May dislike being told what to do (79)
189	Persecutory feelings (17)
189	Seeks little reassurance (49 −)
176	Argumentative (52)
176	May complain about others (69)
172	Influenced by persecutors (6)
168	Meaningless phrases (40)
160	Bizarre thoughts (30)
154	Irrelevant (1 −)
140	Incoherent (45 −)
130	Generally hostile (7)
128	Complains about locked doors (59)
126	Grandiose (13)
123	Some resentment of questions (46)
116	Assertive (22)
88	Stubborn about getting own way (53)

*Cutting line of items arbitrarily determined as identifying the positive end of the type factor is set at 290.

TABLE 2. MSA SCALE RELATING SYMPTOM-CATEGORIES TO TYPE-FACTOR: CATATONIC WITHDRAWN VS. OVER-EXPRESSIVE (COEFFICIENT OF REPRODUCIBILITY = .94)*

Scale Value	Brief Description of Symptom and Item Number
154	Mannerisms or gestures (5)
142	Some positioning (2)
121	Slowed movements (20−)
117	Place disorientation (16−)
93	Sensitive and considerate (24)
92	Undertalkative (62−)
88	No gesticulation (47−)
87	Time disorientation (3)
77	Some fear of doing wrong (37)
45	Very anxious in new situation (38)
44	Seldom looks at E. (57−)
−76	Very cooperative (68)
−84	Optimistic (28)
−89	No restriction in interests (65)
−94	Influenced by persecutors (6)
−103	Bizarre thoughts (30)
−103	Can show initiative (78)
−104	Emotionally labile (12)
−105	Recognizes illness (34)
−108	Initiates greetings to personnel (74)
−122	Overactive (23)
−122	Not perplexed (19−)
−128	Very aware of other patients (77)
−137	Initiates conversation with personnel (70)
−137	Always responds promptly (73)
−139	Not undertalkative (62)
−142	Assertive (22)
−160	Generally hostile (7)
−177	Generally initiates conversation (64)
−206	Overtalkative (4)

*Cutting line arbitrarily set at 100.

TABLE 3. MSA SCALE RELATING SYMPTOM-CATEGORIES TO TYPE-FACTOR: PARANOID (COEFFICIENT OF REPRODUCIBILITY = .78)*

Scale Value	Brief Description of Symptom and Item Number
426	Bizarre thoughts (30)
393	Generally hostile (7)
384	Influenced by persecutors (6)
378	Can show hostility to personnel (71)
368	Can verbalize hostility toward patients (76)
368	Grandiose (13)
366	Overtalkative (4)
355	Stilted speech (39)
343	Concern with abstractions (31)
342	Meaningless phrases (40)
338	Brags (51)
337	Always responds promptly (73)
296	May complain about others (69)
249	Overconcerned about leaving (55)
227	Optimistic (28)
220	Underactive (60−)
197	Not concerned with abstractions (31−)
185	No stilted speech (39−)
169	Expresses despondency (41)
148	Seldom looks at E. (57−)
143	Time disoriented (3)
141	Sloppy appearance (63−)
141	Place disorientation (16−)
113	Moves hands manneristically (42)

*Cutting line arbitrarily set at 360.

TABLE 4. MSA SCALE RELATING SYMPTOM-CATEGORIES TO TYPE-FACTOR: RESISTIVE ISOLATION (COEFFICIENT OF REPRODUCIBILITY = .83)*

Scale Value	Brief Description of Symptom and Item Number
342	May dislike being told what to do (79)
327	Underactive (60−)
320	Social avoidance (66−)
316	Not dependable (80−)
292	Mannerisms or gestures (5)
276	Sensitive and considerate (24)
240	Some fear of doing wrong (37)
218	Very cooperative (68)
215	Recognizes illness (34)
207	Dependable (80)
198	Can verbalize hostility toward patients (76)
196	Some fear of doom (27)
190	Overconcerned about leaving (55)
182	Expresses despondency (41)
177	Initiates conversation with personnel (70)
173	Always responds promptly (73)
168	Time disorientation (3)
160	Generally initiates conversation (64)
155	Some resentment of questions (46)
151	May complain about others (69)
146	Criticism of hospital (56)
142	Irritable (25)
137	Overtalkative (4)
133	Assertive (22)
133	Generally hostile (7)
135	Overactive (23)
116	Stubborn about getting own way (53)

*Cutting line arbitrarily set at 300.

TABLE 5. MSA SCALE RELATING SYMPTOM-CATEGORIES TO TYPE-FACTOR: ANXIOUS-DYSPHORIC (COEFFICIENT OF REPRODUCIBILITY = .90)*

Scale Value	Brief Description of Symptom and Item Number
368	Expresses despondency (41)
342	Recognizes illness (34)
336	Overconcerned about other's opinion (29)
322	Some fear of doom (27)
321	Difficult to make decisions (35)
305	No silly smiling (44−)
278	Sensitive and considerate (24)
264	Residual ability not low (33)
232	Not perplexed (19−)
227	Social avoidance (66−)
218	Brags (51)
207	Stilted speech (39)
207	Opinionated (36)
191	Irrelevant (1−)
184	Generally hostile (7)
183	Generally initiates conversation (64)
176	Overtalkative (4)
170	Complains about locked doors (59)
169	Grandiose (13)
167	Optimistic (28)
163	Can show hostility to personnel (71)
159	Incoherent (45−)
150	Demands excessive attention (67)
148	Silly smiling (44)
142	Meaningless phrases (40)
140	Stubborn about getting own way (53)
134	Assertive (22)
116	Sloppy in appearance (63−)

*Cutting line arbitrarily set at 310.

tained type-factors. The factors accounted for 96.8 per cent of the estimated communality. The factor matrix is omitted to conserve space since there is little value in knowing that a particular individual had the obtained loadings on the factors.[2] It will suffice to know that all factors were well represented by having many heavily loaded subjects for each. Only the second factor was bipolar, and strongly so.

The types are described in Table 1 through 5 in terms of the manifest structure analysis (MSA) relationship of symptom-categories to type-factor loadings. The mean score value for each relevant symptom-category is given in the tables.[3] Thus, a mean of the weighted values for the symptoms encountered in a patient over a particular MSA scale provides an index of his factor loading for that type. The mean scores can be converted to type-factor loading estimates from tables.[2] Attempts to reproduce the factor loadings on a particular type-factor from symptom occurrences for the *Ss* in this study led to the rank-difference correlations of reproducibility found at the heading of each table.

Item identification numbers, followed by a minus in the first five tables, indicate that the items were reversed in direction during scaling (reciprocal categories.[1, p. 145])

Discussion

Before referring to the tables of catescales it will be helpful to consider some of their features that are quite different from tables of factor loadings, thus minimizing likely misinterpretations. Unlike factor loading values, the MSA catescale values are simply points on an arbitrary rank-order scale; they have no zero point and do not represent the degree of variance of the category associated with the criterion. Actually, a category with a high scale value may show less variance associated with the criterion than those categories with mid-range scale values. The only implication of a catescale value is that it is the mean point on a rank-order scale indicating the type-factor loading associated with it. If the catescale value is very high, then individuals with very high type-factor loadings are "plus" scorers for that category. But note that the scaling requirements of minimum scatter (for the intensive model) around the stated scale point result in the rejection of many categories with high scale values because the scatter extending well below the middle range of type-factor scores makes them unreliable indicators of the mean scale value. In practical terms, this means that such a prominent symptom as "suspicious" may be absent from a paranoid scale not because it is usually not encountered in individuals of that factor-type but because it does not discriminate between those in the midrange and those high on the factor-type. This is quite different from factor analysis where the absence of an item from a list of high loadings implies that it is not associated with that factor. Categories not included in the catescales given here may prove to be very good differentiae between catescales, but in the MSA procedure they are selected only because they are good indicators of the magnitude of the particular type-factor for the subjects.

Also, unlike factor loadings, the size of the catescale values are independent of the reliability of the item. While a low factor loading may be obtained for a factorially pure, but unreliable, item, this does not happen in MSA. Throughout the catescale range, items are of comparable reliability, and items of low reliability have automatically been rejected because they will not scale. For this reason, the reliability of any category, representing the degree of type-factor loading, is independent of its catescale position. The big value of MSA lies herein, that we do not have to depend upon a limited number of reliable categories at the top end of the type-factor scale to designate the magnitude of a person's factor loading.

Yet our habit of looking at higher values in factor matrices to determine the nature of the underlying factor is most applicable to MSA. In fact, we must look at

[3]Mean score values for categories given in the tables are mean ranks on the Y axis of the scaling frame multiplied by 10, except for Table 2, where 260 was subtracted from each value to indicate the true zero point.

only the very highest scale values in a catescale to identify it since those below it are lower in scale value because of the intrusion of other factors. They cannot be factorially pure items whose loadings have been lowered merely because of poor reliability, as is often the case with items in factor matrices. It follows that identification of factors is more difficult with catescales than for factor matrices. With these precautions in mind we can turn to a discussion of the catescales obtained.

The *Normal* type-factor seen in Table 1 shows that the five highest (first-listed) items reflect a healthy, flexible adjustment. The hypothetical normal "subject" had a .77 factor loading on this type. As we go down the category scale in Table 1 we note the intrusion of paranoid features to destroy the purity of typing for subjects. It is not surprising that this factor and the *paranoid* type-factor showed a correlation of .43, the largest in Table 6.[2] This factor corresponds closely to the normal factor found in an earlier study [3] of a similar sample of *Ss*.

The *Catatonic Withdrawn* vs. *Over-Expressive* bipolar type-factor seen in Table 2 shows the three highest items to be very prominent symptoms of catatonic stupor. Disorientation is also important. Near the zero point (low scale values) the picture of the sensitive catatonic personality remains but the extreme withdrawal has disappeared and there is a close correspondence to the *anxious-dysphoric* type. Moving down toward the negative end of the scale, some paranoid features appear. Continuing further down, contact with reality becomes hyperacute and hostility and over-expressiveness become marked. The extreme negative end of the scale may represent the excited catatonic condition but the data do not provide clear information on this point. A similar factor appeared in earlier studies. [2, 3]

The *Paranoid* type-factor seen in Table 3 shows the seven highest items to be the outstanding characteristics of paranoid schizophrenics. Egocentricity, hostility, overexpressiveness and delusional thinking are all prominent. Only at the low end of the scale do regressive and withdrawal symptoms appear. The factor appears to represent a fairly pure type at the higher end, and the degree to which a paranoid individual would conform to the type would depend upon the degree to which the paranoid symptoms are present rather than through an intrusion of other type-factor symptoms. Of course, as there is a remission of paranoid symptoms, the individual would revert to the *normal* type, but, more frequently, there is a residual of character problems and he more closely resembles the *anxious-dysphoric* type. A similar factor was found by Guertin [2].

The *Resistive Isolation* type-factor seen in Table 4 shows the four highest items to be a combination of inactivity and a resistance to any attempts to stimulate activity in these individuals. The picture changes very quickly as we go down the scale and the resistiveness is replaced by cooperation and the individuals appear to be sensitive and have insight. At that point, the individuals have many of the features of the *normal* and *anxious-dysphoric* type-factors, so it is not surprising to find that these three types have moderately high intercorrelations. The existence of this factor is confirmed in two earlier studies. [2, 3]

The *Anxious-Dysphoric* type-factor seen in Table 5 shows that the five highest items reflect a combination of insecurity and sensitivity. Anxiety and tension are present which lead to subjective feelings of unhappiness. Conflict and perhaps guilt are present. The high end of the scale discloses no clearly psychotic items and, except for its neurotic flavor, is similar to the *normal* type. The problems appear to be neurotic ones and in our population of psychotic patients this probably represents a remission of psychotic symptoms with the persistence of general inadequacy and conflict. In the middle of the scale, paranoid thought disturbance and hostility intrude to produce a mixture of neurotic and psychotic symptomatology or a pseudoneurotic picture. A similar factor appeared in the earlier study. [3]

Methodological Discussion

The data presented herein are intended to illustrate the methodology. As in most pilot studies, there are many neglects that must be given proper attention in a more thorough investigation. Some of these are: no cross-validation of the type-factors or the MSA catescales, incomplete description of the rating scale used and no demonstration that other raters would view the symptoms in the same way as did this rater, sample restriction to chronic male veterans limiting the general applicability of the types and catescales, as well as many less important weaknesses. Despite these lacks, the empirical findings may prove useful for research purposes with a sample similar to that employed. At any rate, no claim is made as to the establishment of a sound classificatory system. The whole task of differential diagnosis has been ignored in this study, but the methodology proposed and illustrated herein does show promise.

Three of the catescales showed lower coefficients of reproducibility than desired. The lowest value is .78, but upon cross-validation of the catescale it will probably

become apparent that much of the variance in type-factor predicted by the catescale has been lost. There is hope for improving reproducibility and reliability by employing the full five points of the symptom rated, each as a separate category, instead of simply dichotomizing as in this study.

In closing, it would be well to view the unique features of this combined factor analysis—MSA methodology in relation to other typological procedures. It possesses uniqueness of solution that is not characteristic of linkage analysis. The only other method leading to unique typing is Lazarsfeld's[5] latent structure analysis, which leads to the determination of and assignment of individuals to latent classes. It has had limited application because of its computational complexity, but more important, it is applicable only to undimensional problems involving a very few categories and they are customarily dichotomized. In contrast to this, the present methodology is more generally applicable since it provides typing of individuals on multidimensional characteristics, can quickly screen hundreds of categories for relevance to typing, and is capable of employing continuous variables.

Summary

The problems of sampling and describing subjects in studies of schizophrenics are discussed and typology is proposed as a necessary solution. Transposed factor analysis provides a unique solution to typing schizophrenics but there are many difficulties encountered in relating the symptoms to obtained type-factors since they disappear in the process of intercorrelating individuals.

The present study employs manifest structure analysis to establish "catescales" of symptoms, so that the type-factors have a clear description that can be used for classificatory purposes through quantified methods. The combination of factor analysis and manifest structure analysis provides a methodology that can produce quantified classificatory criteria for establishing homogeneous and reproducible samples of schizophrenics.

The five obtained type-factors are: the *normal*, the *catatonic withdrawn* vs. *over expressive*, the *paranoid*, the *resistive isolation*, and the *anxious-dysphoric* types. Sampling limitations of this pilot study are discussed.

References

1. du Mas, F. M. *Manifest structure analysis.* Missoula, Montana: Montana State University Press, 1956.
2. Guertin, W. H. An inverted factor-analytic study of schizophrenics. *J. consult. Psychol.*, 1952, *16*, 371-375.
3. Guertin, W. H. and Jenkins, R. L. A transposed factor analysis of a group of schizophrenic patients. *J. clin. Psychol.*, 1956, *12*, 64-68.
4. Guertin, W. H. and Krugman, A. D. *The Activity Rating Scale.* Gainesville, Fla.: Cooperative Psychological Test Distributors, 1959.
5. Lazarsfeld, P. F. Chapters 10 and 11. In S. A. Stouffer (Ed.). *Studies in social psychology in World War II. Measurement and prediction*, Vol. 4. Princeton: Princeton Univer. Press, 1950.
6. Lorr, M. Multidimensional Scale for Rating Psychiatric Patients (Hospital Form). *Technical Bulletin* TB 10-507 (restricted). Washington, D. C.: Veterans Administration, 1953.
7. McQuitty, L. L. Chapter 4. In B. M. Bass and I. A. Berg (Ed.). *Objective approaches to personality assessment.* Princeton: Van Nostrand, 1959.
8. Rabin, A. I. and King, G. F. Chapter 6 in L. Bellak. *Schizophrenia: A review of the syndrome.* New York: Logos Press, 1958.
9. Wittenborn, J. R. *Wittenborn Psychiatric Rating Scales.* New York: Psychological Corp., 1955.

VALIDITIES OF THREE VISUAL TESTS FOR DIFFERENTIATING ORGANICS FROM SCHIZOPHRENICS AND NORMALS

BERT R. SAPPENFIELD AND ROBERT J. RIPKE
Montana State University *Montana State Hospital*

Problem and Method

This study compared the relative effectiveness of the spiral aftereffect test and two other tests (a visual test and a visual-motor test) for differentiating three groups previously defined as organic, schizophrenic, and normal.

Subjects. The organic *Ss* (21 men and 9 women, patients at Montana State Hospital) had medical diagnoses of chronic brain syndrome, without mental deficiency or psychosis (8 convulsive disorders, 6 with central nervous system syphilis, 3 Korsakoff's syndrome, 3 with cerebral arteriosclerosis, 6 with diseases of unknown or uncertain causes, and one each with senile brain disease, brain trauma, prenatal brain damage, and other cranial infection). The age range was from 22 to 70 years.

The schizophrenic *Ss* (21 men and 9 women, patients at Montana State Hospital) were selected from a complete list of medically diagnosed schizophrenics, so that each *S* was matched with an *S* in the organic group on the variables of sex, age, and educational level.

The normal *Ss* (21 men and 9 women) were hospital employees and other available *Ss* without known organic brain pathology or psychosis. Normal *Ss* were also matched with organic *Ss* on the variables of sex, age, and educational level.

Test procedures. All test stimuli were illuminated by two 60-watt bulbs placed at distances of 1 ft. on either side. Each test was preceded by the general instruction: "This is a special eye test. Keep both eyes open." Each *S* was given tests in the following order: (a) retinal rivalry, trial 1; (b) stereoscopic figure tracing, trial 1; (c) stereoscopic figure tracing, trial 2; (d) retinal rivalry, trial 2; (e) spiral aftereffect, with spirals presented in the order A, B, C.

Spiral aftereffect test. Three 90° spirals mounted on 8-in. discs were rotated at 78 rpm. Spirals A and C appeared to expand while rotating, and induced contracting aftereffects; spiral B appeared to contract while rotating, and induced an expanding aftereffect. *S* was seated 4 ft. from the disc and instructed as follows: "Look at the center of the spiral, and don't take your eyes away until you are told to do so." After the spiral was rotating: "What do you see?" If no answer or an ambiguous answer was given: "Which way is the line moving?" Upon obtaining an answer: "Now watch closely; when I turn the machine off, I want you to tell me what the line does." After 30 seconds, the disc was stopped, and *E* asked: "Now, what is the line doing?" *S* was given a positive score for each trial with each spiral on which he reported perceiving the expected after-effect.

Retinal rivalry test. The apparatus was a Brewster stereoscope mounted vertically above a platform on which the test stimulus was placed. The test stimulus consisted of two 2½ in. x 3¼ in. rectangles mounted on a 4 x 6 card; the left-hand rectangle was red, the right-hand rectangle green; both rectangles were ruled with black lines, spaced approximately 3/8 in. apart, parallel to and including the diagonal from the lower outside corner to the upper inside corner. While *S* looked into the stereoscope, the retinal rivalry stimulus was inserted, with the instruction: "Tell me when the card you are viewing changes." The score consisted of the total number of color and/or line changes reported during two 30-second periods.

Stereoscopic figure-tracing test. The apparatus was the same as for the retinal rivalry test. The test stimulus consisted of a 4 x 6 card, on the left (or right) side of which a cross, a square, and a diamond had previously been drawn, arranged vertically in relation to one another. The cross consisted of two 1-in. perpendicular lines intersecting at the midpoints; the square measured 1 in. on each side; the diagonals of the diamond measured approximately 1 1/4 in. and 3/4 in., respectively. The right (or left) side of the card was blank. The card was inserted under the stereoscope so that the figures appeared on the side of the dominant eye; and, while *S* looked into the stereoscope, instructions were given to "trace over these figures with your pencil." Scoring was performed independently by three judges, using the scoring standards of Terman and Merrill[1] pp. 207-208; 219-221; 230-232); the assigned score for each figure on each trial depended on agreement between at least two of the three judges. Since, possibly due to extremely lenient scoring standards, only 11 *Ss* out of a total of 90 in the three groups failed to achieve passing scores on both trials with the cross, this figure was eliminated from consideration both as a subtest and as a contribution to total score. Scores consisted, then, of the number of successful reproductions, in two trials, of each figure (square and diamond) and of both figures combined.

Results

The data of Table 1 indicate that each of the tests, and each of the subtests, discriminated between the organic group and the normal group, at beyond the .01

Table 1. Comparisons of Organics with Normals on All Variables

Variable	Normal (N = 30)	Organic (N = 30)	χ^2	U	p
Spiral aftereffect					
Spiral A[a]	25	14	7.33		< .01
Spiral B[a]	29	13	17.86		< .001
Spiral C[a]	24	13	7.05		< .01
Total score				187.0	< .00007
Retinal rivalry[b]	29	1	48.60		< .001
				10.0	< .00003
Stereoscopic figure tracing					
Figure B[b]	16	4	9.08		< .01
Figure C[b]	21	9	8.07		< .01
Total score				166.0	< .00003

Note: Entries in second and third columns denote numbers of *Ss* giving successful or above-median performance. All Chi-squares corrected for continuity.

[a]Dichotomized in terms of reporting or not reporting aftereffect.
[b]Dichotomized as near as possible to median of both groups combined.

level of significance. Comparison of the tests and subtests, in terms of their relative discriminating powers, reveals (a) that spiral B (having an expanding aftereffect) discriminated the two groups more successfully than did spirals A and C (having a contracting aftereffect), (b) that the retinal rivalry test discriminated the groups far more successfully than either of the other tests, and (c) that the stereoscopic figure-tracing test compared favorably with the spiral aftereffect test in its ability to discriminate between organics and normals. The total scores for each of the tests distinguished between the normal and organic groups, at extremely low probability levels (.00003 to .00007), as evaluated by the Mann-Whitney U test.

Comparison of the normal group with the organic group (Table 1), in terms of the relative frequencies with which successful or above-median performance occurred, reveals that each spiral of the spiral aftereffect test owed its discriminating ability more to its success in classifying normal *Ss* correctly than to its success in classifying organic *Ss* correctly. The retinal rivalry test, however, was highly successful in classifying correctly the *Ss* in both groups. Figure B of the stereoscopic figure-tracing test was more successful in classifying organics than in classifying normals, while figure C showed moderate, but equivalent, success in classifying the *Ss* in both groups.

The data of Table 2 indicate that two subtests (spirals A and C) of the spiral aftereffect test and both subtests of the stereoscopic figure-tracing test failed to discriminate significantly between the organic and schizophrenic groups, although the total scores for each of the tests, as evaluated by the Mann-Whitney U test, were successful in discriminating between the two groups at acceptable levels of significance. Spiral B of the spiral aftereffect test also discriminated between the schizophrenics and the organics, at an highly acceptable level of significance; again, however, its discriminating ability was attributable mainly to its success in classifying *Ss* in only one of the two groups, the schizophrenic group. The retinal rivaly test, on the other hand, misclassified only two schizophrenics and four organics, on the basis of a cutting score at the median for both groups combined.

Tabular data are omitted for the comparison of schizophrenics with normals, since schizophrenic *Ss* performed at levels approximating the performance of normal

TABLE 2. COMPARISONS OF ORGANICS WITH SCHIZOPHRENICS ON ALL VARIABLES

Variable	Schiz. (N = 30)	Organic (N = 30)	χ^2	U	p
Spiral aftereffect					
Spiral A[a]	21	14	2.47		>.10<.20
Spiral B[a]	28	13	15.10		<.001
Spiral C[a]	21	13	3.33		>.05<.10
Total score				240.0	<.001
Retinal rivalry[b]	28	4	35.42		<.001
				41.0	<.00003
Stereoscopic figure tracing					
Figure B[b]	21	14	2.47		>.10<.20
Figure C[b]	16	9	2.47		>.10.<.20
Total score				280.0	<.01

Note: Entries in second and third columns denote numbers of *Ss* giving successful or above-median performance. All Chi-squares corrected for continuity.

[a]Dichotomized in terms of reporting or not reporting aftereffect.
[b]Dichotomized as near as possible to median of both groups combined.

Ss on both the spiral aftereffect test and the stereoscopic figure-tracing test, both in terms of subtest scores and in terms of total scores. The only test which discriminated significantly between the schizophrenic and normal groups was the retinal rivalry test, on which 19 normals and 10 schizophrenics achieved above-median scores, when the median was based on the scores of the two groups combined. The resulting Chi-square value was 4.27 (significant at the .05 level), while the value of U was 248.5 (significant at the .002 level).

SUMMARY

This study investigated the relative effectiveness of the spiral aftereffect test, a retinal rivalry test, and a stereoscopic figure-tracing test for discriminating between groups previously defined as organic, schizophrenic, and normal.

Each of the tests, and each of the subtests, discriminated between organics and normals at acceptable levels of significance. Total scores on each of the tests, but only one subtest (spiral aftereffect involving perception of expansion), also discriminated at acceptable levels of significance between organics and schizophrenics. The only test which discriminated significantly between schizophrenics and normals was the retinal rivalry test. The retinal rivalry test, as compared with the other two tests, was found to classify correctly much larger proportions of *Ss* in the different groups.

REFERENCE

1. TERMAN, L. M. and MERRILL, MAUD A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.

THE USE OF DRAWINGS OF HANDS AS A PROJECTIVE MEDIUM FOR DIFFERENTIATING NORMALS AND SCHIZOPHRENICS

EDWIN E. WAGNER

University of Akron

Problem and Method

Recent work in the prediction of anti-social behavior has suggested that drawings of human hands may serve as useful projective devices for assessing personality.[1] This exploratory study was concerned with the possibility of differentiating a "known" clinical group, schizophrenia, from a normal population utilizing this new projective approach.

The subjects were 50 diagnosed male schizophrenics randomly selected from Massilon State Hospital, Massilon, Ohio, and 50 male college students from Akron University. The mean age of the schizophrenic group was 42.5, S. D. 17.6. The normal group had a mean age of 20, S. D. 1.3. The schizophrenics were diagnosed as 10 chronics; 8 undifferentiated; 6 simple; 5 catatonic; 7 hebephrenic; 20 paranoid.

The stimuli consisted of 10 cards approximately 4 by 5 inches. On each of these cards except the last a hand was drawn in a recognizable but ambiguous pose. Each subject was asked to "tell what the hand is doing". The last card is blank and the subjects were asked to "imagine a hand doing something". The responses were recorded verbatim by the examiner. The responses were analyzed and scored as follows:

1. An *interpersonal* response: another person is explicitly or implicitly involved in the action, *e.g.*, "shaking hands", "holding up traffic", "fighting someone", "talking", "dancing".

2. An *inactive* response: hand is resting or posed, little or no energy is exerted with attendant lack of interpersonal and/or environmental accomplishment, *e.g.*, "hand resting on a table", "holding fingernails up to dry", "sleeping".

3. An *active* response: movement of a non-interpersonal nature, *e.g.*, "picking up a coin", "reaching for something", "throwing a ball".

4. A *maladjustive* response: tension, pain, disfigurement or despair is projected into the response, *e.g.*, "all clenched up", "crippled", "got sores on it", "drowning", "paralyzed", "dead", "hanging from a ledge".

5. A *withdrawal* response: an inability to respond meaningfully to the card which presumedly indicates some degree of pathological withdrawal from interpersonal and/or environmental situations. Withdrawal responses can be subdivided into Description, Failure and Bizarre responses.

   a. *Description*: the presence of the hand is acknowledged without imparting a meaningful action tendency, *e.g.*, "it's a hand", "a left hand", "a hand with five fingers".

[1] As reported by Z. A. Piotrowski at the International Congress of Psychology in Bonn, Germany, and reviewed in *Science Newsletter* for August 13, 1960, page 109.

Table 1. Distribution of Responses, Significance Levels and Correlational Values for the Normal and Schizophrenic Groups on Each Scoring Category

Number of Responses per Protocol	Number of Individuals in Each Group Producing a Given Number of Responses Per Protocol for Each Scoring Category									
	Interpersonal		Inactive		Active		Maladjustive		Withdrawal	
	Norm.	Schiz.	Norm.	Schiz.	Norm.	Schiz.	Norm.	Schiz.	Norm.	Schiz.
0	2	24	14	13	0	19	18	36	36	5
1	4	8	11	3	6	9	22	5	12	3
2	7	9	7	7	9	7	7	3	2	11
3	22	6	11	7	10	6	2	1	0	6
4	10	2	5	9	12	7	1	3	0	6
5	3	0	0	5	9	0	0	1	0	4
6	2	1	1	3	4	0	0	0	0	5
7	0	0	1	2	0	0	0	1	0	1
8	0	0	0	0	0	2	0	0	0	3
9	0	0	0	1	0	0	0	0	0	2
10	0	0	0	0	0	0	0	0	0	4
Total	50	50	50	50	50	50	50	50	50	50
Median Test	<.001		>.05		<.001		<.001		<.001	
Phi Values	.56				.40		.36		.81	

b. *Failure*: subject is unable to respond at all to a particular drawing.

c. *Bizarre*: a response completely out of keeping with the drawn contours of the hand, *e.g.*, "a big black bug creeping along the wall", "the hand of God with the tree of life growing out of it".

The above categories were predetermined by the author because of what seemed to be their logical *a priori* projective connotations and the simplicity of scoring. As a control, the protocols were administered and scored by a graduate student, Captain Ford Smart, U. S. A. F., who had no knowledge of the interpretive significance of the scoring categories. For each subject, a tally was made of the number of responses produced in each category. Median tests for significance and Phi tests of correlation were then computed between the experimental and control groups for each major scoring category. Since there was an obvious disparity between the ages of the two groups, median tests were also computed to establish the significance, if any, between age and each diagnostic scoring category for the schizophrenic group. This was not done for the normal group because of the similarity in ages.

Results and Discussion

There were no significant differences on the median test between age and scoring categories for the schizophrenic groups suggesting that any differences between the experimental and control group were attributable to factors other than age. Table 1 summarizes the distribution of responses on each scoring category for the schizophrenic and normal groups with significance levels and correlational values. All scoring categories were significant beyond the .001 level with the exception of the *inactive* category. Maximum differentiation was obtained from the *withdrawal* category which produced a correlation of .81 between the presence of *withdrawal* responses and schizophrenia.

Apparently, the use of hand drawings as a projective medium for differential diagnosis should be pursued further on an experimental basis since, notwithstanding the obvious limitations of this study, significant differentiation has been obtained between the normal and schizophrenic groups. If the connotations of the scoring categories are taken literally, a picture of the schizophrenic emerges which is in keeping with what is known about their overt behavior: reduction in interpersonal relationships, lowering of activity level, maladjustive behavior, and withdrawal from reality situations. The average administration time in this experiment was only about five to ten minutes per subject.

Summary

Significant differentiation was obtained between a normal and schizophrenic group using drawings of hands as an experimental projective device.

SHORT FORMS OF THE WISC FOR USE WITH EMOTIONALLY DISTURBED CHILDREN[1]

RICHARD ENBURG[2], VINTON N. ROWLEY, AND BETH STONE

State University of Iowa

In a clinical setting, the psychologist frequently is faced with the problem of obtaining the maximum amount of information in a minimum amount of time. Often information is desired regarding the intellectual functioning of the child as well as his personality dynamics. Particularly with the emotionally disturbed child, a question about his intellectual ability may arise, as the presenting problem frequently centers on his inability to adjust in the school situation. This raises the problem of determining whether the child is having academic difficulty associated with intellectual factors, emotional factors or both. If consideration of time enters into the selection of testing materials, the psychologist must select a relatively brief test of intellectual functioning in order to devote more time to personality evaluation.

[1]The authors are indebted to the staff of the Massachusetts Institute of Technology Computation Center and International Business Machines for their assistance in processing the data for this study.

[2]Now at University of Connecticut.

The Wechsler Intelligence Scale for Children (WISC) is particularly adaptable to this type of problem, as the clinician may select only particular subtests for administration. However, when only selected subtests are administered, the reliability of the prorated quotient is questionable. Several approaches have been used to determine the reliability of such shortened batteries. One method is to determine reliability as a function of the correlation between the subtests utilized and the Full Scale intelligence quotient. Bridges[2] published a nomograph based upon the intercorrelations between the subtests and the intelligence quotient of Wechsler's standardization sample. His graphs furnish estimates of the reliability of various combinations of subtests. As he mentions, the graphs are applicable only to the population on which the scale was standardized.

Another approach to developing short forms of the WISC involves preselecting certain combinations of subtests and determining the correlations with the Full Scale quotient. Carleton and Stacey[3] obtained correlations ranging from .64 to .88 for 21 such combinations. Their sample consisted of children evaluated for possible mental deficiency, without organic impairment. The applicability of these short forms for emotionally disturbed children is somewhat questionable in light of the findings of Yalowitz and Armstrong[13]. Utilizing various combinations of five subtests, they obtained correlations in the order of .55 to .61 between the abbreviated scale and the complete examination of children referred to a child guidance clinic. Another investigation by Simpson and Bridges[9] found that the sum of Vocabulary and Block Design correlated .874 with the total weighted score for child guidance cases. In the above studies the short forms investigated have been selected either on the basis of the original correlations obtained by Wechsler or the investigator's preference for certain subtests—either of which may have introduced bias.

An alternative approach to evaluating intelligence is to employ some instrument which is shorter or more convenient to administer than the entire WISC, yet which correlates well with the Full Scale quotient. Among the tests investigated have been the Kent EGY, Raven's Progressive Matrices (RPM), Raven's Colored Progressive Matrices (RCPM), the SRA Primary Mental Abilities Test, the California Test of Mental Maturity (CTMM), and the Columbia Mental Maturity Scale (CMMS). Barratt[1] found that the RPM correlated .754 with the WISC Full Scale quotient while the CMMS had a correlation of .606. On this basis he proposed that the RPM be used as an alternative to the WISC rather than the CMMS. Martin and Wiechers[7] found that the RCPM correlated .91 with the Full Scale quotient of the WISC. Both of the above studies were based on a sample of school children. Stacey and Carleton[10] obtained a correlation of .55 between the RPCM and the WISC Full Scale quotient with children evaluated for possible mental deficiency. Stemple[11], using children of superior social-economic level, found a correlation of .58 between the WISC Full Scale quotient and the SRA quotient. Delp[5] obtained a correlation of .618 between the WISC Full Scale quotient and the Kent EGY. He concluded that this scale should not be utilized as an evaluative instrument in preference to the subtests of the WISC as they had better correlations with the Full Scale quotient than the Kent EGY.

In the above studies the subjects were not classifiable as being emotionally disturbed children. Since the degree of comparability among emotionally disturbed children, school children and mentally deficient children is unknown, it should not be assumed that findings based on one group can be generalized to another. In addition there is the problem of relative efficiency. If some combination of subtests from the WISC can be found which correlates as well or better than the above alternative tests, while not requiring more time, it would appear that the shortened WISC being more familiar to the clinician, would be a more appropriate instrument. The purpose of the present investigation was to derive short forms of the WISC such that (a) administration time could be lessened without appreciably reducing the reliability, and (b) previous sources of possible bias from preselection of subtests or utilization of Wechsler's original sample would be eliminated.

Method

The population for this study consisted of children referred to the Child Psychiatry Service, Psychopathic Hospital, State University of Iowa, during the years 1957 to 1960. All cases were selected to meet the criteria of being without a primary diagnosis of organic involvement or mental deficiency (defined as having a Full Scale quotient of less than 70 on the WISC) and with some form of emotional difficulty. The mean age was 11.29 years, range 5-7 to 15-11; both males (N = 118) and females (N = 27) were included for a total sample of 145 cases. All Full Scale quotients were computed on the basis of ten subscales, with Mazes and Digit Span being omitted.

The basic information sought was a series of correlations between various combinations of the subtests of the WISC and the total weighted score. Since one consideration for developing such a shortened battery was a minimization of time involved in administration, it was decided to evaluate only combinations of three, four and five subtests. In order to avoid any investigator preference in regard to the selection of the various combinations, it was decided to form all the possible different combinations of subtests for a given size of the short form. Thus for a short form which includes only three subtests, there are 120 possible different combinations. For the four subtests form there are 210 possible combinations, and for the five subtests form 252 combinations are possible. The above information involved forming the sum of the scaled scores for the various possible combinations and then correlating these with the total scaled sum for the entire battery, resulting in 582 correlations.

In utilizing the above procedure, two possible factors which might bias the results were avoided. The first is the assumption that the correlations between the subtests and the Full Scale quotient would be the same for a population of emotionally disturbed children as was found for Wechsler's standardization sample. Secondly, preferential selection of combinations of subtests is avoided.

Results

The mean Full Scale IQ of the present sample was 97.18 (SD = 15.41), the mean Verbal Scale IQ was 97.26 (SD = 16.08), and the mean Performance Scale IQ was 97.61 (SD = 15.54). The differences between the means were obviously not significant. The sum of the Verbal scaled scores correlated .90 with the sum of the Full

Table 1. Correlations Between Combinations of Three Subtests and the Sum of Full Scale Scaled Scores

Subtests	*R*
I, C, BD	.93
C, A, BD	.92
I, V, BD	.92
I, S, BD	.92
A, S, BD	.92
S, V, BD	.92
C, V, BD	.91
C, S, BD	.91
I, V, OA	.91
A, V, BD	.91

scaled scores, while the sum of the Performance scaled scores correlated .85 with the sum of the Full scaled scores. These results are comparable to those reported by Wechsler [12] for his standardization sample.

The correlations of three, four and five subtests ranged in value from .77 (PC, OA, Cod) to .96 (C, A, S, BD, Cod). It is not surprising that when the number of subtests was increased the correlated values obtained were also enhanced. For example, PC, OA and Cod correlated .77, but with the addition of BD the correlation

TABLE 2. CORRELATIONS BETWEEN COMBINATIONS OF FOUR SUBTESTS AND THE SUM OF FULL SCALE SCALED SCORES

Subtests	*R*
C, A, S, BD	.94
I, V, PC, OA	.94
I, C, PA, BD	.94
I, V, PC, BD	.94
I, S, PA, BD	.94
I, C, BD, Cod	.94
I, C, PC, BD	.94
I, C, A, BD	.94
I, V, PA, BD	.94
I, C, S, BD	.94

rose to .82. When PA was added to this combination of four subtests the correlation increased to .85.

The ten combinations of three, four and five subtests which correlated most highly with the sum of the weighted scores are presented in Tables 1, 2 and 3. Only ten combinations are listed for each size of the shortened test, since it is probable that they will be of most interest to the clinician.

TABLE 3. CORRELATIONS BETWEEN COMBINATIONS OF FIVE SUBTESTS AND THE SUM OF FULL SCALE SCALED SCORES

Subtests	*R*
C, A, S, BD, Cod	.96
C, A, S, PA, BD	.96
I, C, V, PC, OA	.96
I, C, A, PC, BD	.96
I, C, V, PC, BD	.95
I, C, S, BD, Cod	.95
I, C, A, PA, BD	.95
C, A, V, PC, BD	.95
I, C, V, BD, OA	.95
I, C, S, PA, BD	.95

DISCUSSION

The results indicate that a few subtests of the WISC may be used to assess intellectual functioning without greatly reducing the reliability of the obtained quotient. In this respect, the results are comparable to those reported by Carleton and Stacey [3] and Simpson and Bridges [9], but are in contrast to those obtained by Yalowitz and Armstrong [13]. This latter discrepancy may reflect sampling differences since their sample also included organically impaired and defective children. In addition, Yalowitz and Armstrong mentioned that their population showed considerable intra-test scatter and a larger standard deviation on the Performance Scale than on the Verbal Scale, whereas the present study did not reveal similar variability.

One possible objection to the use of any short form of the WISC is that the clinician loses diagnostic information regarding the child's functioning on those subtests which are omitted. Although this information may be desirable in some instances, recent investigations by Cohen [4] and Rowley [8] cast some doubt on the validity of utilizing differential performance on the subtests as a diagnostic procedure. The present findings lend credence to the use of a few subtests to evaluate intellectual functioning of emotionally disturbed children. They do not, however, justify the use of a short form in cases where the child presents problems of organic involvement and or mental retardation. In such cases, the results of Carleton and Stacey [3] are more aptly applied.

SUMMARY

The purpose of the present investigation was to derive short forms of the WISC for use with emotionally disturbed children, such that administration time could be lessened without appreciably reducing reliability and previous sources of possible bias from preselection of subtests or the utilization of Wechsler's original sample would be eliminated. All combinations of three, four and five subtests of the WISC were correlated with the sum of the scaled scores for the Full Scale of 145 children whose diagnosis was emotional disturbance. It was concluded that selected subtest combinations could reliably estimate the intellectual functioning of emotionally disturbed children.

REFERENCES

1. BARRATT, E. S. The relationship of the Progressive Matrices (1938) and the Columbia Mental Maturity Scale to the WISC. *J. consult. Psychol.*, 1956, *20*, 294-296.
2. BRIDGES, C. Nomographs for computing the "validity" of WISC or Wechsler-Bellevue short forms. *J. consult. Psychol.*, 1959, *23*, 453-454.
3. CARLETON, F. O. and STACEY, C. L. Evaluation of selected short forms of the Wechsler Intelligence Scale for Children. *J. clin. Psychol.*, 1954, *10*, 258-261.
4. COHEN, J. The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *J. consult. Psychol.*, 1959, *23*, 285-299.
5. DELP, H. A. Correlations between the Kent EGY and the Wechsler batteries. *J. clin. Psychol.*, 1953, *9*, 73-75.
6. LITTELL, W. M. The Wechsler Intelligence Scale for Children: Review of a decade of research. *Psychol. Bull.*, 1960, *57*, 132-156.
7. MARTIN, A. W. and WIECHERS, J. E. Raven's Colored Progressive Matrices and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.*, 1954, *18*, 143-144.
8. ROWLEY, V. N. Analysis of the WISC performance of brain damaged and emotionally disturbed children. *J. consult. Psychol.*, in press.
9. SIMPSON, W. H. and BRIDGES, C. C., JR. A short form of the Wechsler Intelligence Scale for Children. *J. clin. Psychol.*, 1959, *15*, 424.
10. STACEY, C. L. and CARLETON, F. O. The relationship between Raven's Colored Progressive Matrices and two tests of general intelligence. *J. clin. Psychol.*, 1955, *11*, 84-85.
11. STEMPEL, ELLEN F. The WISC and the SRA Primary Mental Abilities Test. *Child Develpm.*, 1953, *24*, 257-261.
12. WECHSLER, D. and WEIDER, A. Tests of intelligence: C. Wechsler Intelligence Scale for Children. In A. Weider (Ed.), *Contributions toward medical psychology*. New York: Ronald Press, 1953, Pp. 522-529.
13. YALOWITZ, J. M. and ARMSTRONG, RENATE G. Validity of short forms of the Wechsler Intelligence Scale for Children (WISC). *J. clin. Psychol.*, 1955, *11*, 275-277.

THE CEILING OF THE NEW STANFORD-BINET

WALLACE A. KENNEDY, HAROLD MOON, WILLARD NELSON AND RONALD LINDNER, JACK TURNER

Florida State University

PROBLEM

The low ceiling of the 1937 Revision of the Stanford-Binet has been one of the major limitations encountered in its use with above-average adolescents and adults.[1, 2] In spite of the provision of an average adult and three superior adult levels above the fourteen-year level, it happens disturbingly often that subjects pass items at the upper limit of the test. The Third Revision of the Stanford-Binet Intelligence Test[5] presents some improvements: (a) The extension of IQ tables to chronological age 18; (b) the selection of the more discriminating items from both previous forms and combining them into one new form; and (c) the use of standard scores at each of the chronological age levels, with a mean of 100 and an SD of 16.

Method

In the Summer of 1960, the authors had an opportunity to put the upper level of the 1960 Binet to a reasonable test, repeating, with minor modifications, a study[3] of mathematically gifted adolescents. Twenty-nine eleventh-grade students participated in a Summer Mathematics Institute sponsored by Florida State University and the National Science Foundation. The program, which has been described by Nichols[4], is in its third year. Students were selected for the program on the basis of mathematical ability as determined by scores on the Eleventh Annual Mathematical Contest sponsored jointly by the Mathematical Association of America and the Society of Actuaries. For the most part the students were the top 29 entries; an exception was made to obtain the intended minimum of 6 girls. In two instances girls were selected instead of male competitors with higher scores. All but 3 of the 29 selected had reached their sixteenth birthdate; these three were aged 15-0, 15-6, and 15-9.

The Sequential Test of Educational Progress, Mathematics Form, was administered upon arrival of the students at the University as a verification of their selection test scores. The range on the STEP Test was from 62 to 98; mean, 85.9; SD, 8.4. In addition to the STEP Test, two further measures were obtained: teacher rankings based upon subjective evaluation, and a ranking from examination scores in two subject areas—Theory, and Sets and Functions. The correlation between the the STEP Test were .79 for Theory rankings and .77 for Sets and Functions rankings.

In the previous study[3], the WAIS and WISC rather than the 1937 Revision of the Binet were employed due to the upper limit restrictions imposed by the latter when administered to superior adolescents. Because of the age level differences of the 1959 Math Campers, both WISC and WAIS were administered despite their doubtful comparability. Consequently five of the fifteen-year olds were administered both instruments as a reliability check and the reliability scores were found to be within the standard error of the test. Therefore, in the present study, in an attempt to evaluate the intended improvements of the 1960 Revision of the Stanford-Binet, the WAIS was administered in conjunction with Forms 1A and 1B of the STEP Test and the 1960 Revision of the Binet. It was hypothesized that the latter should have an extended upper limit.

Results

Concerning the ceiling of the new S-B, it was found that all of the students passed some items at all age levels; that is, a maximal level could not be obtained on these students. Thus one of the original objections to the Binet still seems to be applicable.

In spite of this objection, however, the concurrent validity of the new S-B appears to be high (Table 1). It correlates .52 with the Full Scale WAIS, .50 with the

Table 1. Means, SD's and Correlations Between Measures

Measure	Mean	S. D.	WAIS Full	WAIS Verb.	WAIS Perf.	WAIS Total	STEP A	STEP B	STEP S-B
WAIS									
Full	133.8	6.6				.43*			
Verbal	136.8	9.2				.45**			
Performance	125.1	10.6				.07			
STEP									
Total	85.9	8.4							
A	43.8	4.5							
B	42.2	4.6					.68**		
S-B (1960)	137.4	8.2	.52**	.61**	−.03	.50**	.52**	.41*	
Theory			.34						.48**
Sets and Functions			.44*						.53**

*$p < .05$. **$p < .01$.

total STEP, .48 with the Theory rating, and .53 with the Sets and Functions ratings. These correlations were significant at the .01 level of confidence. The WAIS Verbal correlated .61 with the S-B while the Performance correlated only −.03. While it has been pointed out [2, p. 197] that the Wechsler Performance Scale fails to provide a correlation as high as the Verbal Scale with the S-B, it is interesting to note that in the present study no significant correlations were found between the Wechsler Performance Scale and any of the other measures obtained. Indeed, none of the correlations with this scale were above .12. But the results were very different with the Verbal Scale. The mean of the current sample on the Verbal Scale was 136.8, with an SD of 9.2. Thus the mean of 137.4 on the new S-B is very close to the mean on the Verbal Scale of the WAIS.

Summary

It appears that improvements have been made in ceiling characteristics of the 1960 Stanford-Binet Test, although inadequacies in ceiling still exist. There have been improvements in derived IQ's due to selection of items and to the extension of the chronological ages in the tables. As a research instrument for the study of superior adolescents, the new S-B seems to render effective, though not ideal, measurement at the top of the scale.

References

1. Anastasi, Anne. *Psychological testing.* New York: The Macmillan Company, 1959.
2. Cronbach, Lee J. *Essentials of psychological testing.* 2nd ed. New York: Harper & Brothers, 1960.
3. Kennedy, Wallace A. and the Human Development Clinic Staff. A multi-dimensional study of mathematically gifted adolescents. *Child Develpm.*, December, 1960 (in press).
4. Nichols, Eugene A. Summer mathematics camp for talented high school students. *The Bulletin of the National Association of Secondary School Principals*, 1959, *43*, 100-103.
5. Terman, Lewis and Merrill, Maud. *Stanford-Binet intelligence scale.* Boston: Houghton Mifflin Company, 1960.
6. Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation, 1955.

HANDEDNESS AND DIGIT SYMBOL PERFORMANCE

RICHARD J. BONIER AND CHARLES HANLEY

Michigan State University

Problem

The format of the Digit Symbol Test, used by Wechsler in his WAIS, WISC, and Wechsler-Bellevue measures, may handicap a certain type of left-handed subject. Procedures in this speeded test require that subjects, working from left to right across the page, enter code symbols below numbers arranged in horizontal rows. A model containing the code appears directly above the answer section of the page.

Right-handed individuals typically hold a pencil so that the hand lies below and to the side of the implement. This "conventional" posture is also found among left-handers. Some left-handed persons, however, grasp the pencil in a "crabbed" fashion, positioning it between hand and writer. As the hand and arm move from left to right in this style of writing, they mask what has just been written and block the view of material above the line. With the Digit Symbol Test, the hand and arm in the crabbed style cover both previously encoded material and the model giving the correct symbol-number pairs. A conventional writing technique does not impose this potential handicap. The present study investigates the effect of these two different writing styles upon Digit Symbol performance.

In his instructions for WISC administration of the Digit Symbol Test (Coding B), Wechsler[1] indicates that examiners should "Place a second record form before the (left-handed) Subject folded so the relevant code is showing, since a left-handed Subject often covers the code when writing on the record form." This modification of directions is absent in the manuals for the WAIS and the Wechsler-Bellevue. In the present study, comparisons are made between Digit Symbol Performance of "crabbed" and "conventional" writers using both standard and modified instructions for test administration.

Standard Instructions

Subjects. Because the "crabbed" technique is relatively rare in women, comparisons are restricted to males, students in introductory psychology classes at Michigan State University. Of these, 37 were right-handed, 14 conventional left-handed, and 15 crabbed left-handed.

Procedure. Subjects were tested individually in two sessions. In the first, each was given a special "unbiased" revision of the Digit Symbol Test employing the code used for Form I of the Wechsler-Bellevue. The format of the test was changed from horizontal to vertical alignment; that is, subjects worked down rather than across the page. On the form for right-handers, the model containing the number-symbol code was placed on the left-hand side of the page with the answer section to the right, while the numbers to be given symbols appeared in the left of each pair of boxes. These arrangements were reversed in the form for left-handers. Both forms are "unbiased", because obstruction from the hand and arm in viewing the model and past work is considerably reduced for crabbed left-handers and absent for the other subjects. Five to eight weeks subsequent to the initial session, subjects returned to take the Digit Symbol Test in its common Wechsler-Bellevue form (Form I). Standard instructions were used in both this and the preceding session.

Results. Means and standard deviations for the three groups on both forms of the test appear in Table 1. Analysis of covariance holding constant individual differences on the "unbiased" version yields an F-ratio of 8.62 (df 2, 62), significant beyond the .005 level, for differences on the standard version of the Digit Symbol Test. An analysis of variance of scores on the "unbiased" test gives an F-ratio of 1.86, not statistically significant. The "within-groups" correlation between scores on the two sessions is .86.

Table 1. Digit Symbol Test Performance of Right-Handed, Standard Left-Handed, and Crabbed Left-Handed Subjects under Standard and Modified Instructions

Test Form		Standard Instructions RH (N = 37)	SLH (N = 14)	CLH (N = 15)	Modified Instructions SLH (N = 12)	CLH (N = 13
Unbiased:	Mean	56.2	52.9	51.9	49.8	51.6
	SD	8.95	6.44	6.45	8.11	5.81
W-B:	Mean	62.6	61.0	54.3	60.1	55.5
	SD	8.55	6.79	8.28	7.99	7.56

These findings permit several conclusions. To begin with, the high correlation between forms indicates that the "unbiased" version measures much of what is tapped by the Wechsler-Bellevue form. Second, left-handers and right-handers are not appreciably different in coding ability as such; witness their performances on the "unbiased" form. Finally, the format of the Wechsler-Bellevue form handicaps crabbed left-handers. (The difference between crabbed and conventional writers on the W-B version of the test is large enough to yield significant differences without the covariance analysis; analysis of variance of performance on the second session gives an F-ratio of 5.69, significant at the .01 level.)

Modified Instructions

Subjects. The 25 left-handed subjects, 12 conventional and 13 crabbed, were males in introductory psychology classes at MSU in the term following the above study.

Procedure. The procedure was the same as in the preceding investigation, except that on the second of the testing sessions the crabbed left-handers received the modified instructions quoted earlier from the WISC Manual.

Results. Means and standard deviations of scores on both forms of the Digit Symbol Test appear in Table 1. Analysis of covariance of performance on the W-B form, holding constant individual differences appearing on the "unbiased" form, gives an F-ratio of 7.64 (df 1, 22), significant beyond the .025 level. The "within-groups" correlation between the two forms is .73.

Table 1 gives very similar means for comparable groups. The difference between conventional left-handers in the two studies is statistically insignificant using analysis of covariance; the same is true for the scores of crabbed left-handers, a further indication of the failure of the modified instructions to remove the handicap.

Discussion

Consideration of a subject's handedness as a factor in intelligence-test evaluation is apparently a fairly rare phenomenon. The fact that crabbed left-handers are significantly handicapped by the format of Wechsler's version of the Digit Symbol Test indicates that they should, when responding to it as part of a full scale IQ test, be expected to obtain commensurately lower IQ scores. This assumption is warranted, of course, only to the degree to which it is safe to regard results when the Digit Symbol Test is given alone as comparable to results obtained for it in the course of a complete intelligence test examination.

We cannot explain the results by making the assumption that crabbed left-handers are basically less intelligent than others, because this factor has been controlled statistically by use of "unbiased" pre-test scores and analysis of covariance. Unfortunately, an examiner who wishes to adjust the score of a crabbed left-hander to compensate for the bias has no normative material on which to base such a correction.

The study indicates that the modified instructions offered by Wechsler do not remove the handicap imposed by the format of the test. Few of the crabbed left-handers seemed to pay attention to the altered instructions, and a certain amount of confusion was apparent in those who did. Whether these instructions are more successful in the complete testing situation is an open question. The "unbiased" form of the test employed in the present study appears to overcome the handicap. This revision, however, is less practicable than Wechsler's Digit Symbol Test because separate forms are required for right- and left-handers.

These conclusions are based on results with college students. Perhaps they may be extrapolated throughout the entire range of intellectual ability. Two other alternatives, however, can be entertained; i.e., crabbed left-handers of lower intelligence may be handicapped less, or handicapped more, than is the case with the college students. Which of the three is correct is a matter for further investigation.

Summary

A special version of the Digit Symbol Test (Form I, Wechsler-Bellevue), arranged vertically so as to abolish any advantage stemming from right-handedness, was given to 37 right-handed and 14 left-handed subjects with "conventional" writing technique, and to 15 left-handers who use a "crabbed" technique. Five to eight weeks later, the subjects were retested with the Wechsler-Bellevue form. Crabbed left-handers were significantly poorer in performance on this second session.

Groups of 12 conventional and 13 crabbed left-handers were used in a study of the effectiveness of modified test instructions suggested by Wechsler. Again, results

indicate significantly lower performance for the crabbed left-handers on the standard form of the test.

The "within-groups" correlations between scores on the first and second sessions of .86 in the initial study and .73 in the second, demonstrate that the special and standard versions of the Digit Symbol Test tap similar abilities.

That differences between crabbed and conventional groups on the standard form of the test are due to writing technique and not to coding ability *per se* is shown by the use of analysis of covariance employing scores on the special version of the measure as the matching variable. In addition, when subjects are compared in performance on the special version, no significant differences appear. The format of the standard version of the Digit Symbol Test, it appears from these findings, imposes a handicap specific to left-handers who write with a crabbed style of penmanship.

Reference

1. Wechsler, D. *Wechsler Intelligence Scale for Children, Manual.* New York: Psychological Corporation, 1949.

A VALIDITY STUDY OF TAMKIN'S "MMPI SCALE MEASURING SEVERITY OF PSYCHOPATHOLOGY"[1]

DANIEL R. LUTZKER

Fort Dix, N. J.

Problem

In 1959 Tamkin [3] presented some preliminary data on an MMPI scale measuring the severity of psychopathology (P) consisting of the F scale and the non-overlapping items of Grayson's [2] Critical Item (CI) scale. For convenience, the non-overlapping items of the CI scale are called X in this paper. Tamkin [3] reports that a sample of 50 psychotics scored significantly higher than a sample of 25 neurotics and personality disorders. No data are given for normal subjects. This paper presents data regarding the performance of normal subjects, tests the ability of the scale to discriminate between groups differing in degree of psychopathology, and provides information concerning the relative usefulness of the P, F and X scales in distinguishing between such groups.

Method

MMPI answer sheets were obtained for five groups of subjects. Basic trainees at Fort Dix who were in their first (N = 28), fourth (N = 34) and eighth (N = 31) week of training served as the normal control groups. These subjects had been tested as part of a research project [1] on the effects of basic training. Because of the stress encountered in such training it was expected that measures of psychopathology would show a steady increase as training progressed. All three groups of normal basic trainees, however, were expected to score significantly lower than the experimental groups. The experimental subjects were first court-martial offenders (N = 45) and 25 consecutive patients who were administered the MMPI for diagnostic purposes after January 1, 1959. These groups were selected so as to include only male basic trainees. The group of patients contained both in-patients and out-patients and the subjects varied considerably in their symptomatology.

[1]The author gratefully acknowledges the assistance of Stanley G. Fullwood and Christian Kristensen in the statistical treatment of the data.

The MMPI answer sheets were scored for all of the clinical and validity scales, the P and X scales, and for mean T-score of the clinical scales. The last of these measures was used as a criterion of degree of pathology.

Results

The results in Table 1 show that for all three scales there is an increase in mean score as psychopathology increases and that a wide gap exists between the three normal and the two abnormal groups. This suggests that all three scales measure psychopathology. However, the size of the S.D.'s indicates severely limited usefulness for individual prediction as Tamkin found.

Table 1. Means and S. D.'s on the F, X and P MMPI Scales for Five Groups

Group	N	F		X		P	
		Mean	S. D.	Mean	S. D.	Mean	S. D.
NBT 1	28	4.07	2.69	1.43	1.99	5.50	4.02
NBT 4	34	5.15	4.52	2.00	2.61	7.15	6.74
NBT 8	31	7.06	5.39	2.71	2.62	9.77	7.10
FCM	45	13.90	12.17	4.70	3.97	18.60	15.83
Patients	25	14.80	10.71	6.60	2.04	21.40	14.98

The mean P score for normal basic trainees in the eighth week of training falls about half way between those of the two pathological groups studied by Tamkin. This may reflect increased anxiety due to situational stress. An alternative hypothesis is that some of the subjects may have been feigning pathology in the hope of obtaining a discharge from the Army.

A further check was made on the validity of the F, X and P scales. Each scale was correlated with the mean T score (based on the nine clinical scales) for the subjects in the three normal groups. The resulting correlations were: between mean T and F, .84; between mean T and X, .65; and between mean T and P, .87. Using mean T as the criterion of psychopathology, all three scales are shown to have some validity. However, P does not appear to correlate better with mean T score than does F and therefore the value of P is dubious. This conclusion is supported by the fact that both the F and P scales discriminate all of the normal groups from the two pathological groups at the .01 level but fail to discriminate between the two pathological groups. F and P can be used interchangeably for all practical purposes and there is little to be gained by scoring answer sheets for P when F is already available.

Summary

Validity data, both predictive and correlational, are presented on Tamkin's Psychopathology (P) scale. While the results indicate that the scale is valid, its validity is no greater than that of the F scale. Therefore it does not appear worthwhile to score MMPI answer sheets for this new scale when a conventional scale is equally useful.

References

1. Ekman, P. Psychological reactions to basic training. Mimeographed report, Mental Hygiene Consultation Service, Fort Dix, New Jersey.
2. Grayson, H. M. *An Admissions-testing Manual.* Mimeographed report, Veterans Administration Neuropsychiatric Hospital, Los Angeles, California.
3. Tamkin, A. S. An MMPI scale measuring severity of psychopathology. *J. clin. Psychol.*, 1959, *15*, 56.

AN ABBREVIATED QUALITATIVE SCORE FOR THE PORTEUS MAZES

JOHN R. BARRY AND LOUIS EVERSTINE
Western Psychiatric Institute and Clinic, University of Pittsburgh
AND
JOHN P. KLEMAN
Psychological Research Center, University of Hawaii

While scoring the qualitative errors on the Porteus Maze Test[1], it was noted that two subscores, Crossed (or touched) Lines, and Lift Pencil, each contributed larger portions of the variance to the Qualitative (Q) Score than did the other six subscores. The variances based upon data from a sample of 70 white, female, hospitalized mental patients were: Lift Pencil 69.56; Crossed Lines 34.57; First Third Error 17.81; Wavy Lines 16.32; other subscore variances were smaller. It is proposed that the sum of the two subscores having the largest variances, Lift Pencil and Crossed Lines, be used in place of the Q Score for ordering patients along the qualitative performance dimension.

In two sub-samples, Pearson *r*'s between the Q Score and the sum of the two subscores were: .91 (N = 40) and .90 (N = 30). The sub-samples consisted of white females, aged 20 to 48, tested at the time of their admission to a university psychiatric hospital. From the size of these *r*'s, it appears that the sum of these two subscores can be used to order the patients in approximately the same way as the Q Score, with a saving in scoring time and less chance for error.

This observation was tested using data from two other independent samples. One consisted of 42 normal Chinese boys and 57 normal Chinese girls in the ninth grades of three Honolulu public schools. The other group consisted of 49 boys and 40 girls tested at the Hawaii Boys and Girls Training School, a state institution for delinquents. Pearson *r*'s between the Q Score and the sum of the Lift Pencil and Crossed Lines subscores were .86 for the Chinese group, and .93 for the delinquent group.

The consistency of these relationships between the Q Score and the subscores, manifest in three different samples, supports the generality of the finding. Therefore it is proposed that the sum of the two subscores, Lift Pencil and Crossed Lines, may serve as an adequate substitute for the more complex Q Score, when ranking or ordering of the subjects is the goal.

Reference

1. Porteus, S. D. *The Maze Test and Clinical Psychology.* Palo Alto, California: Pacific Books, 1959.

THE APPLICABILITY OF AN AMERICAN DELINQUENCY SCALE TO JAPANESE SUBJECTS

T. SHIRASA AND T. AZUMA[1]
Hokkaido University

Problem

In 1958, Quay and Peterson[2] reported a brief scale for the differentiation of delinquent and nondelinquent males. In several cross-validational samples of American boys, approximately 67 per cent of the subjects (Ss) were correctly identified by the measure. The applicability of such scales to other populations, however, is always open to question, and the present study was undertaken to determine whether similar differentiation could be achieved between delinquent and nondelinquent adolescents in Japan.

Subjects and Procedures

The scale was translated into Japanese and administered to a total of 357 Ss. Forty of these were in a training school for delinquents at Engaru in Hokkaido. Another delinquent sample was gathered by testing 130 Ss in a corrective school at Chitose. But it seems to be some difference between two these samples in a degree of delinquency proneness. On the other hand, 187 Ss were chosen, for nondelinquent sample, from public secondary and high schools at Sapporo.

Results

The major findings of the study are presented in Table 1.

A cutting score between 6 and 7, as suggested in the original report[2], leads to proper identification of 88 per cent of the training school Ss at the expense of 26 per cent incorrectly classified Ss among the public school pupils. Boys in the corrective school displayed intermediate scores. Their mean score was 8.11, as compared with 11.75 for Ss in the training school, and 4.99 for nondelinquents.

A split-half reliability of .81 was found for the training school Ss. Separate reliabilities were computed for primary and secondary corrective

[1]We wish to thank the administrators of the schools for permission to test the subjects, and Donald R. Peterson of the University of Illinois for help in preparing the manuscript.

TABLE 1. CUMULATIVE INCIDENCE OF VARIOUS SCORES FOR THREE GROUPS OF SUBJECTS

Score	Training School N	%	Corrective School N	%	Public School N	%
15 & above	12	30	10	7	3	2
14 " "	12	30	19	15	4	2
13 " "	15	38	21	16	5	3
12 " "	17	43	28	22	11	6
11 " "	20	50	35	27	13	7
10 " "	23	58	42	32	23	12
9 " "	26	65	52	40	31	16
8 " "	31	78	60	46	43	25
7 " "	35	88	73	56	53	26
6 " "	38	95	84	65	69	34
5 " "	40	100	98	75	91	46
4 " "			107	82	113	58
3 " "			115	88	130	67

school boys and odd-even coefficients of .71 and .75 were found. A Kuder-Richardson coefficient for public school Ss was .75. The scale was readministered after a two-week interval to students in the high school, with test-retest correlation of .85.

DISCUSSION

The discriminative power of the scale in our sample is actually as high as in the groups on which it was originally developed. The correspondence between assumed degrees of socialization or delinquency proneness and scores on the measure is also surprisingly close, considering all the differences between the present Ss, and those examined in deriving the scale. But the findings go beyond a demonstration of validity; they offer a fairly important theoretical implication.

Recently Gough[1] proposed that certain concepts and dimensions, such as socialization, responsibility, etc., represent cultural universals, i.e., they are applicable to all societies at all times and in all places. As a start in testing this proposition, the Socialization (So) scale of the California Psychological Inventory was translated into Hindi, Punjabi, Spanish, and Italian, and administered to various samples in India, Costa Rica, and Italy. Delinquent and nondelinquent Ss in all these societies were differentiated about as sharply as in the United States, and the applicability of the So scale to rather widely different cultural groups was quite impressively established[1].

Quay and Peterson's delinquency scale[2] correlates about −.70 with the So scale, and can probably be regarded as another measure of socialization. The present results thus offer additional confirmation of Gough's contention regarding the cultural generality of the dimension, and vitiate still further the usual pessimism about allegedly invalidating "cultural drifts" and "moving criteria" in empirically developed scales such as these. The extent to which a person has incorporated cultural norms is probably important in any society, no matter what the particular content of those norms might be. There is now increasingly good reason to suppose that existing measures of socialization transcend linguistic, racial, and cultural limitations to a degree never dreamed of when the scales were first developed. All this opens up some rather exciting possibilities for cross-cultural research on personality, quite beyond any practical descriptive or predictive utility the scales might have.

SUMMARY

The cultural generality of an American delinquency scale was examined by administering it to delinquent and nondelinquent boys in Japan. The discriminative efficiency and reliability of the scale were approximately as high as in the United States. Since the scale can be regarded as a measure of socialization, the demonstrated generality of that dimension has now been extended beyond the United States and three other societies, to another culture in the Orient.

REFERENCES

1. GOUGH, H. G. Cross-cultural studies of the socialization continuum. Paper read at APA meeting, Chicago, Illinois, 1957.
2. QUAY, H. C. and PETERSON, D. R. A brief scale for juvenile delinquency. *J. clin. Psychol.*, 1958, *14*, 139-142.

VISUOCONSTRUCTIVE AND VERBAL RESPONSES IN CHRONIC BRAIN-DAMAGED PATIENTS AND FAMILIAL RETARDATES*

EARL E. BALTHAZAR, RONALD E. TODD, DON H. MORRISON

Caro (Mich.) State Hospital for Epileptics

and

PETER W. ZIEBELL

Lapeer (Mich.) State Home and Training School

Introduction

This study reports a joint investigation by the respective psychology departments of a hospital for epileptics and a home and training school for mental retardates into differential diagnosis of chronic brain syndrome and familial mental retardation. In the past, a number of studies involving organic populations have reflected the imprecision of the term "brain-damage" by failing to take into account significant behavioral differences resulting from localized areas of neurologic abnormality. Neurologic studies[5, 8] relating functioning behavior to locus of cerebral injury or abnormality have shown, however, that speech difficulty and visuo-constructive or perceptuomotor disabilities seem to stem from certain unilateral lesions or cortical malfunctions in the left and right hemispheres.

Anderson[1], Stacey and Levine[12], Klφve[7], and Reitan[9, 10] established that an IQ difference between verbal and performance scales has value in differentiating left-right and bilateral brain damage in that damage to the left hemisphere was associated with lowered verbal scale scores, damage to the right hemisphere with lowered performance scores, and diffuse or bilateral damage with slightly depressed scores on both verbal and performance scales. Balthazar and Morrison[2] reported that a difference of seven IQ points or greater between Wechsler performance and verbal items was of diagnostic value in determining predominant left-right and indeterminate unilateralized brain damage.

Hunt[6] found that familial retardates were superior to brain-injured defectives on most visuoconstructive tasks. Gallagher[4] compared matched brain-injured and familial groups and found those with organic brain syndrome to be inferior to the familial group in visuoconstructive or perceptuomotor ability, in integrating verbal concepts, and in making verbal associations. Sievers[11] reported familial cases to equal or exceed the language level of the brain-injured. No controls were exercised in these studies regarding locus of cerebral abnormality and functioning behavior.

This study tests the hypothesis that minor visuoconstructive abnormalities underlie interhemispheric involvement by comparing known brain-damaged cases, each having predetermined unilateral abnormalities, with cases classified as so-called familial or undifferentiated retardates. We predicted specifically that organics with predominant left-right and indeterminately unilateralized hemispheric abnormalities would tend to be consistently lower on Block Design, Object Assembly, and Digit Symbol subtests than those classified as familial retardates. The study will also examine verbal ability in the two groups.

Method

Subjects were 90 chronic organic patients at the Caro State Hospital for Epileptics who were administered Form II of the Wechsler-Bellevue Scales as part of routine admission procedure in the standard manner[13]. Criteria for the selection of organic subjects were that subjects have (a) unequivocal evidence in their medical histories, EEG's, and general medical examinations of organic brain syndrome; (b) parents and siblings of estimated average intelligence; (c) no gross neurologic or

*The authors are grateful to C. David Hellman for interpreting the electroencephalograms.

sensory abnormalities; and (d) valid Wechslers indicating defective intelligence. An eight-channel Grassi, Model 3 AR, ink-writing electroencephalograph with ten electrodes was used to make bipolar recordings and to determine, along with Wechsler scale discrepancies, hemispheric localization of brain damage. Each patient was given two or more EEG's.

Group R, on the basis of the above criteria, consisted of 30 patients diagnosed as having predominant right-sided brain damage. Group N included 30 patients with a diagnosis of indeterminately unilateralized or bilateral brain damage. Group L was composed of 30 patients classified as having predominant left-sided brain damage. The term "predominant" served to qualify classifications of right-sided and left-sided lateralization because no rigid controls existed for minimal diffuse or bilateral damage which could have caused mild disturbance in the contralateral hemisphere.

An additional group of 30 subjects, Group F, from the Lapeer State Home and Training School was also administered Form II of the Wechsler Scales. The criteria for the selection of Group F were: (a) all were mentally retarded on the basis of the Wechsler; (b) there was unequivocal evidence of mental retardation among their parents and siblings; (c) there was no evidence of trauma contributing to mental retardation; and (d) there was no evidence of chronic or acute brain syndrome in the medical histories or examinations. Group F, therefore, consisted of individuals classified in a state home and training school as familial or undifferentiated retardates. All organic groups, Groups R, N, and L, as well as the non-organic Group F, were equated on the basis of full scale IQ, age, and sex. Subtest means and standard deviations were calculated for each of the groups and the subtest mean differences were compared and tested for significance by the *t* test.

Results

There were significant weighted score differences between Group R (predominantly right-sided brain-damaged), and Group F, (familial), on the Information, Arithmetic, and Similarities subtests of the verbal scale, and Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol subtests of the performance scale. Group R was clearly superior to Group F on verbal test items but not on the performance scale. Group F achieved significantly higher weighted scores on all performance test items including visuoconstructive tasks.

TABLE 1. SCORES, MEAN DIFFERENCES AND LEVELS OF SIGNIFICANCE BETWEEN W-B SUBTESTS FOR FAMILIAL (F), RIGHT HEMISPHERE ORGANICS (R), INDETERMINATE ORGANICS (N), AND LEFT HEMISPHERE ORGANICS (L)

Subtests	F	R	N	L	F − R		F − N		F − L	
					$M_{diff.}$	t	$M_{diff.}$	t	$M_{diff.}$	t
Inf.	3.27	5.20	4.60	3.33	1.93	.01	1.33	.001	.06	N.S.
Comp.	4.93	5.40	4.93	4.23	.47	N.S.	.00	N.S.	− .70	N.S.
D.S.	4.13	5.30	4.65	4.09	1.17	N.S.	.52	N.S.	− .04	N.S.
Arith.	2.57	4.60	3.83	2.90	2.03	.01	1.26	.05	.33	N.S.
Sim.	4.43	6.00	5.13	4.37	1.57	.05	.70	N.S.	− .06	N.S.
Voc.	5.07	5.67	5.10	4.20	.60	N.S.	.03	N.S.	− .87	N.S.
P.A.	6.63	4.10	5.70	5.33	−2.53	.01	− .93	N.S.	−1.30	N.S.
P.C.	6.50	4.63	5.57	6.86	−1.87	.01	− .93	N.S.	.36	N.S.
B.D.	6.43	4.20	5.13	6.33	−2.23	.01	−1.30	N.S.	− .10	N.S.
O.A.	10.47	4.93	7.27	8.57	−5.54	.001	−3.20	.01	−1.90	.01
D.Sy.	5.80	3.20	3.97	4.67	−2.60	.001	−1.83	.001	−1.13	.05
Verb.Scale	65.30	73.43	70.53	64.50	8.13	.02	5.23	.02	− .80	N.S.
Perf. Scale	78.10	59.70	70.73	77.30	−18.40	.001	−7.37	.01	− .80	N.S.

When Group N, those with bilateral or diffuse brain damage, was compared to Group F, verbal superiority, favoring the organic group, was reduced to the Informa-

tion and Arithmetic subtests. Group F achieved significant scores in the Object Assembly and Digit Symbol subtests.

There were no statistically significant differences in verbal scale subtests when Group L, the predominantly left-sided organics, was compared with Group F. The previously noted trend in favor of the organics did not occur in the left-sided group. Group F maintained superiority in Object Assembly and Digit Symbol subtests and both groups attained their highest scores in these two subtests.

Discussion

The right-sided group was unique in showing significantly lower scores on Block Design. Reitan[10] in an unpublished monograph reported individual examinations indicating impairment on Block Design to be associated with right hemispheric involvement. Present findings, however, regarding brain-damaged groups, R, N, L and the familial group F, with the notable exception of the Block Design subtest, tend to support the hypothesis that basic differences between these organic groups and familial retardates seem to lie in visuoconstructive responses to a degree independent of hemispheric unilateralization. In this, our findings are in gross agreement with previous psychological studies regarding perceptuomotor differences. They are not, however, wholly consistent with results obtained in the investigations of Gallagher [4] and Sievers[11] regarding verbal factors in differentiating organic from so-called undifferentiated groups.

Using institutional samples, the right-sided group showed superiority over familials in verbal scale items. This was less true of indeterminately unilateralized cases and not true of predominant left-sided cases. Wechsler verbal scale test items are perhaps more general in nature than the more specialized language tests of Gallagher [4] and especially Sievers[11]. It seems difficult, however, despite the possible limitations of omnibus testing, to believe that an organic group whose cortical abnormalities are, for the most part, confined to the minor hemisphere should be verbally inferior to a group whose language difficulties seemed sufficient for institutional commitment.

While the findings suggest behavioral differentiation of measured abilities in the two hemispheres, inquiry is necessary to clarify further probability of more precise localization. This is a question we cannot answer with our present data. The question of localization is a major one, since the inaccuracy of free hand placement of electrodes, the superficiality of cortical contact limiting hemispheric localization, and sampling error play an essential part. Reitan[6], in a rigidly controlled study, was forced to a somewhat similar conclusion regarding the occurrence of vascular changes and edema in the hemisphere.

Hécaen, Penfield, Bertrand, and Malmo[5] in a ten-year study using precise neurologic controls concluded that visuoconstructive disabilities are integral parts of both hemispheres. They reported also that disturbances of spatial orientation, which we believe to underlie visuoconstructive disability, occurred especially from right parietotemporal lesions. This would seem consistent with our results that visuoconstructive difficulty occurs with left hemispheric involvement, or with malfunction in both hemispheres, but is more especially marked in the right hemisphere.

Present findings imply that visuoconstructive responses may have some diagnostic value in discriminating chronic cases of organic brain syndrome from the undifferentiated, subcultural familial type in terms of measured abilities. Such responses may also underlie prior considerations of hemispheric localization to the extent that they may be basic to both hemispheres and with Wechsler scale discrepancies may detect unilateral lesions or malfunctions independently of supplementary neurological information. To investigate this possibility, however, additional study must undertake to determine whether psychometric study can accurately detect behavioral manifestations of inter-hemispheric symptoms and whether detection of these symptoms validly discriminates additional diagnostic groups.

Summary

The use of Form II of the Wechsler-Bellevue Scales to discriminate 90 organic brain-damaged patients and 30 institutionalized familial retardates was investigated. Results showed that those with predominant right-sided brain damage were superior to familials on verbal items and inferior on all performance tasks, including visuoconstructive test items. Closer score approximation was achieved when diffuse brain-damaged cases were compared to familials, but brain-damaged cases maintained some verbal superiority as well as performance inferiority on visuoconstructive tasks. The exception was Block Design. When predominant left-sided organics were the comparison group, however, verbal superiority was lost. With the exception of Block Design, the familial group maintained superiority in visuoconstructive responses. Implications of these findings were discussed in terms of the basic hypothesis, and consideration given to further study to determine if visuoconstructive responses, combined with Wechsler scale discrepancies, are essential in detecting hemispheric unilateralization independently of neurologic criteria.

Bibliography

1. Anderson, A. L. The effect of laterality localization of focal brain lesions on the Wechsler-Bellevue Subtests. *J. clin. Psychol.*, 1951, *7*, 149-153.
2. Balthazar, E., and Morrison, D. The use of Wechsler-Bellevue intelligence scales as diagnostic indicators of predominant left-right and indeterminate brain damage. In press.
3. Fi-her, G. M. Differences in WAIS verbal and performance IQ's in various diagnostic groups of mental retardates. *Amer. J. ment. Defic.*, 1960, *65*, 256-264.
4. Gallagher, J. J. A comparison of brain-injured and non-brain-injured mentally retarded children on several psychological variables. *Monogr. Soc. Res. Sci.*, 1947, *65*, 329-333.
5. Hecaen, H., Penfield, W., Bertrand, C., and Malmo, R. The syndrome of apractognosia due to lesions of the minor cerebral hemisphere. *A. M. A. Archiv. Neurol. Psychiat.* 1956, *75*, 400-433.
6. Hunt, Betty. Performance of mentally deficient brain-injured children and mentally deficient familial children on construction from patterns. *Amer. J. ment. Defic.*, 1959, *63*, 679-687.
7. Klφve, H. The relationship of differential EEG patterns to the distribution of Wechsler-Bellevue scores. *Neurology*, 1959, *9*, 871-876.
8. Penfield, W., and Roberts, L. *Speech and Brain Mechanisms.* Princeton, N. J.: Princeton Univ. Press, 1959, 198-205.
9. Reitan, R. M. Certain differential effects of left and right cerebral lesions in adult humans. *J. comp. physiol. Psychol.*, 1955, *48*, 474-477.
10. Reitan, R. M. The effects of brain lesions on adaptive abilities in human beings. Unpublished monograph, Indianapolis, Ind., Indiana University Medical Center, 1959.
11. Sievers, Dorothy. A study to compare the performance of brain-injured and non-brain-injured mentally retarded children on the differential language facility test. *Amer. J. ment. Defic.*, 1959, *63*, 839-847.
12. Stacey, C. L., and Levine, J. Correlational analysis of scores of subnormal subjects on the Stanford-Binet and WISC. *Amer. J. ment. Defic.*, 1953, *57*, 481-483.
13. Wechsler, David. *The Wechsler-Bellevue Intelligence Scale Form II: Manual for Administering and Scoring the Test.* New York: The Psychological Corporation, 1946, 37-64.

STANDARDIZATION OF THE GRAPHOSCOPIC SCALE: THE CONTENT OF CHILDREN'S DRAWING*

JUSTIN PIKUNAS AND HUGH CARBERRY[1]

University of Detroit

PROBLEM

This study investigated some relations between developmental level and content of children's drawings on the basis of partially structured perceptual cues. The majority of studies in the literature dealing with the content of children's drawings have been longitudinal studies. Eng's[1] detailed study of the drawings of her niece, Hildreth's[4] investigation of more than four thousand drawings made by a boy between the ages of two and eleven years old, and McCurdy's[6] study of more than one thousand drawings made by a boy between the ages of two and one-half and seven years, are major sources. The leading cross-sectional study conducted on the content of children's drawing was McCarty's[5] analysis of 31,239 drawings of children between the ages 4 and 8 years.

METHOD

The instrument used in collecting the data for this study was the Pikunas Graphoscopic Scale (PGS)[7, 8], which is a semistructured projective drawing test primarily designed for children. The PGS contains ten drawing plates, eight of which have semistructured perceptual cues, while one is fairly structured but is compensated by another which is free drawing of the "object you most often draw," and the eleventh is a written answer to the question, "What else would you like to draw?" The fact that the PGS is a semistructured test adds to the interest of this study of the content of children's drawings for, up until this time, the majority of studies has been centered around either free or particular drawings[3, 8].

Our standardization sample consisted of 350 children from the Detroit area schools, who ranged in age from five to fifteen years and who provided approximately 3,200 individual drawings.

The focus of this statistical content analysis will be on the following three areas: (a) a comparison of the four most frequently drawn content categories at the various ages, (b) a comparison of the content drawn at three major developmental levels; namely, middle childhood, late childhood, and early adolescence, and (c) an analysis of the specific tendencies depicted for most perceptual cues of the PGS.

RESULTS

Table 1 compares the four most frequently drawn content categories at the various ages and reveals that the overall chi square value of 151.87 is significant at the .01 level. The chi square values obtained for each of the four major content categories were also significant at the .01 level. Hence we can conclude that the various ages did differ significantly in the frequency of the content drawn. In order to make the frequencies more meaningful they have been reported in terms of percent drawn at each age.

Table 1 also shows that the most frequently drawn objects for the group as a whole were those involving the human person. In the "human total" category (which is a summary percentage including total human persons, human heads, halves of human bodies and human details), a definite trend was noted in that there was a gradually decreasing percentage of human content as the chronological age increased. This finding at the earlier age levels (five through eight years) is in agreement with McCarty[5] who interpreted this trend as an increasing interest in other areas. The human form drawn at all age levels was usually an adult figure. An increase is again

*This study was in part presented at the 1960 convention of the American Psychological Association, Chicago, Illinois.

[1]Now at Villanova University.

TABLE 1. PERCENT OF THE FOUR MOST FREQUENTLY DRAWN CONTENT CATEGORIES AT THE VARIOUS AGE LEVELS

Ages	N	Human	Factory made Objects	Animals	Houses
5	31	51.0%	26.7%	9.8%	12.3%
6	25	41.8	40.0	9.4	8.6
7	31	35.5	40.0	9.2	15.0
8	39	36.9	34.7	11.8	16.3
9	34	40.2	30.6	18.1	11.0
10	33	25.0	45.2	12.8	16.6
11	33	35.4	34.7	14.0	15.7
12	34	45.1	32.6	12.5	9.6
13	30	35.7	31.3	25.0	7.7
14	46	28.0	48.0	16.5	7.4
15	13	45.2	31.6	15.7	7.3
Drawings N	3074	1140	1131	433	370
χ^2		63.18*	155.65*	71.41*	60.24*

*Significant at the .01 level.
Overall chi square for Table 1 is 151.87, significant at the .01 level.

noted at the eleven through thirteen ages. However, this does not exceed the percentage at the lowest age level and may be interpreted as one of the consequences of the increased social awakening typical of these years.

The second most frequently drawn type of objects for the group as a whole was "factory-made objects". In this category a fairly steady increase with age occurs, with the lowest percent being drawn by the five year age group. The peak percentage for this category was reached at the ten and fourteen year age groups. The "factory-made objects" category, due to its inclusive definition, also yielded more specific trends within the different age levels, ranging from simple household objects and depictions of toys at the earlier age levels, to the depiction of moving objects, like trains, cars and boats at the older age levels. The drawing of dresses and hats by the girls at the older age level was also revealed, pointing to the possibility of sex role awareness and acceptance.

The third most frequently drawn object was found in the animal category. Generally, the percentage of animal content drawn fluctuates from year to year. The peak of animal drawing was reached at the thirteen year level, while the lowest percentages of animal content were drawn at the five, six, and seven year level. The lack of emphasis on animals at the younger ages is in direct contrast to the young child's tendency to identify with animals. This is probably due to the emphasis placed on human figure drawing in the school and home. While animal content was found to be consistently lower than human content at all ages, the manner of depicting animals at the various ages differed considerably. At the earlier ages there is very little distinction between the animal drawings and the drawings of the human figure; the distinction is largely made by the child's labeling of the content. As the age level increased, a change in orientation appeared, with the animal's body usually placed in a horizontal position, and with detail appearing in terms of appendages, such as tails, whiskers and fur. The rather large increase in percent at thirteen years may be related to an expression of the increasing instinctual urges which are concomitant with puberty, and also as a rise of interest in pets.

Houses were the fourth most frequently drawn objects, but the only discernible trend revealed was the rather steady decrease in percent in the older age groups (eleven through fifteen). This decrease may be the consequence of the child's increasing maturity and his attempt to depict more complex objects. A general improvement in the symmetry, neatness, and number of details in the houses was also a product of increasing age. At the upper age levels, the drawing of a house was usually accompanied by surrounding details such as flowers, trees and grass, which

made the portrayal more of a scene than the isolated drawings of a house that appeared at the earlier age levels. There was greater tendency to draw buildings other than houses as the age level increased. Furthermore, content dealing with nature, landscapes, designs, abstract material and plant life were among the more frequent. Drawings of food, usually of the sweet variety, emerge at seven and are fairly dominant up to eleven. The striving for completion and composition was evidenced in all the objects drawn and was a definite accompaniment of increasing mental growth.

Developmental levels. In an attempt to obtain a clearer picture of the growth trends, an analysis of the content was made of the four major categories at three basic developmental levels: middle childhood (7 and 8 year group), late childhood (10 and 11 year group), and early adolescence (13 and 14 year group). The results given in Table 2 show that the overall chi square (48.53) is highly significant at the .01 level. From this it may be concluded that the groups differed significantly in terms of the content drawn, although the chi square values obtained for "human" and "factory-made objects" did not reach the required level of significance. These categories are almost equally popular at all three developmental levels.

TABLE 2. PERCENT OF THE FOUR MOST FREQUENTLY DRAWN CONTENT CATEGORIES AT THREE MAJOR DEVELOPMENTAL LEVELS

Content	Middle Childhood	Late Childhood	Early Adolescence	Total Frequency	x^2
Human	35.8%	30.0%	31.2%	615	2.71
Factory-made objects	37.6	40.1	41.0	753	1.96
Animals	10.6	13.4	20.1	281	23.08*
Houses	15.8	16.2	7.5	251	9.69*
Total Drawn	622	632	646	1900	

The overall chi square for Table 2 is 48.53, significant at the .01 level.
*Significant at the .01 level.

In the "animal" category, which shows significant differences, there is a steady increase in percent drawn from level to level. Again, there is evidence for the expansion of interest due to life experiences. It has been suggested that the sharp rise in percent at early adolescence may be interpreted as an expression of the vitalized instinctual drives typical of the early adolescent level of development [9].

In the "house" category, there is a slight increase in percent in late childhood followed by a decrease in early adolescence. In early adolescence the drawing of houses is apparently replaced in part by more complex architectural objects. Emancipation from the home occurring at this age is possibly a factor of high significance.

Item Findings. In order to see the usual and popular, an analysis of the content drawn for each PGS item (plate) was also made. The percentages mentioned in this analysis are based on the total number of drawings at the various ages. The major trends for each PGS plate number are noted in the subsequent summary, except the first which is too structured to have relevance for any content analysis.

The most frequently drawn objects for Item 2 were animal figures; while the second and third most frequently drawn objects were the human figure and factory-made objects, respectively. At the lower age levels (5, 6 and 7 years) the percentage of animals drawn (25.8, 38.3, 19.4) is lower than the human content (34.4, 34.4, 41.5), but then it gradually rises, reaching a peak at 13 years (70.8). The highest percentage of human content is reached at seven years (41.5), and then there is a gradual decline reaching the lowest frequency at 11 years (18.6) and then rising again at 12 years (34.2). In the frequencies of factory-made objects no trend is indicated.

The most frequently drawn object for Item 3 was the human category while the second and third most frequently drawn objects were animals and factory-made objects, respectively. The trend revealed in human category is essentially the same

as in Item 2 where the highest percentage was in the early years (63.3, 70.3, 54.0) with a gradual decline followed by a rise at the 12 year level (55.8). The lowest percentage of animal content is found at years 5 (3.3) and 6 (3.7) followed by an intermittent rise and fall in percentages at the other age levels. The highest percentages of animals drawn are found at years 8 (34.1) and 13 (30.0). The highest percentage of factory-made objects was at the 13 year level (20.0). No trend is indicated for this category of content.

The most frequently drawn object for Item 4 was human, followed by factory-made objects and animals. The highest percentage of human content was at the 8 year level (67.5) and the lowest at the 5 year level (42.7). The highest percentages of factory objects were found at the 5 (35.7) and 10 year level (34.4), while the highest percentage of animals drawn is found at the 13 year level (30.0). For either category of content no age trend is indicated.

The highest frequency for Item 5 was factory-made objects followed by human drawings and food. None of the content categories reveal any age trend. The highest percentages of drawings in this category were found at the 12 (57.1) and 14 year levels (65.9). The lowest percentages of human content drawn were the 12 (35.6) and 14 year levels (24.9), while the highest percentages for food category were found at the 13 (17.2) and 11 year levels (13.7).

The most frequently drawn objects for Item 6 were factory-made objects followed by human content and plants, respectively, with no trend indicated on either content. The highest percentage of factory-made objects was found at 12 years (67.8) and the lowest at 5 (37.9), while the highest percentage for plant category was found at the 5 year level (20.6).

The most frequently drawn objects for Item 7 were factory-made objects followed by aircraft and nature. A gradual increase in factory percentages is noticed reaching a peak at the 10 year level (67.6) and dropping only slightly thereafter. The lowest percentage in this category is found at the 5 year level (23.3). The highest percentage was found at the 6 year level (32.1) for aircraft category while the lowest was at the 12 year level (30.0). The highest percentage for nature category was found at the 8 year level (31.4). In the last two content categories no age trends are shown.

The most frequently drawn objects for Item 8 were factory-made objects, closely followed by houses and then human content. The highest percentages for factory objects are found at the 8, 9 and 10 year old levels (39.4, 41.1, 45.4). The highest percentages were found at the 11 and 12 year levels (58.8, 58.8) for houses while the lowest percentage was found at the 5 year level (7.1). The highest percentage of human content was at the 6 year level (20.4). No trend is discernible in either content category.

The most frequently drawn object for Item 9 was factory-made objects followed by human content and houses, respectively. In factory-made objects no trend is indicated. The highest percentages were found at the 12 (44.7) and the 14 year level (47.7). Without a trend, the highest percentages of human content were found at the 8 and 11 year level (46.9, 45.8). In house drawing a gradual rise is seen reaching a peak at 10 years (20.5) and then declining.

The most frequently drawn object for Item 10 was factory-made objects followed closely by human content and houses. In factory objects and houses no trend is indicated. The highest percentage for the first category was found at the 10 year level (38.2) while for the houses it was found at the 5 year level (50.0). In human content there is a gradual decrease in percentages up to the 11 year where a slight increase is noted. The highest percentage of this category was the 5 year level (32.3) and the lowest at the 9 (8.5) and 10 year levels (8.7).

The most frequently mentioned objects for Item 11 were human followed by houses and factory-made objects. From the 6 year level on, there is a gradual decrease in human percentages to the age of 10 followed by a gradual increase in the later years (11, 12 and 13). The highest percentage of human content is found at the 13 year level (35.7). The high for houses category is found at the 5 year level (67.0)

and again at the 8 year level (50.0). The highest percentages for factory objects were found at the 7 (32.2) and 10 year levels (32.3) while the lowest percentages were found at the 5 year level (5.8).

Summary and Conclusions

A total of 350 Detroit area children between the ages of five and fifteen provided approximately 3,200 drawings on the PGS plates. Tabulations and analyses of the drawing content result in the following major findings:

1. The group as a whole tended most frequently to draw human individuals, factory-made objects, animals and houses, respectively. Significant differences for each of the four major categories were found at all ages. Using the chi square technique, differences were found to be significant at the .01 level. In the "human" content category there was a general decrease in the percent of human content drawn as the chronological age increased. The increment in the drawing of "human" content at the later ages (12 and 15) was interpreted as one consequence of increased social awareness. In the "factory-made objects" category an increasing percent was noted as the age increased, with a peak being reached at the ten and fourteen year age groups. In the "animal" category no discernible trend was noted but a relatively high peak was reached at thirteen years. This high peak was interpreted as an expression of the early adolescents' magnifying instinctual urges. A decreasing percent of "houses" drawn in the eleven through fifteen age brackets was the only discernible trend noted for the "house" category.

2. The stimuli of the PGS were found to be rather well adapted for eliciting a great variety of content at all ages. The most recurring categories other than the above mentioned were as follows: nature, landscapes, designs, abstract material, plant life and food.

3. A comparison of the content drawn at middle childhood, late childhood and early adolescence revealed significant frequency differences in the "animal" and "factory-made objects" categories. In the "animal" category a gradually increasing percent of drawings was noted from earlier to later developmental levels.

4. In an analysis of each particular item of the PGS, specific trends common to that item were also identified.

5. A striving for completeness, composition, improvement in orientation toward test cues, and enlarged content and its detail production were noted as the chronological age increased.

References

1. Eng, Helga K. *The psychology of children's drawing.* New York: Harcourt, Brace, 1931. (First published in German, 1927.)
2. Goodenough, Florence L. Studies in the psychology of children's drawings. *Psychol. Bull.*, 1928, *25*, 272-283.
3. Goodenough, Florence L. and Harris, D. B. Studies in the psychology of children's drawings: II. 1928-1949. *Psychol. Bull.*, 1950, *47*, 369-433.
4. Hildreth, Gertrude. *The child mind in evolution: a study of developmental sequences in drawing.* New York: King's Crown Press, 1941.
5. McCarty, S. A. *Children's drawings.* Baltimore: Williams and Wilkins, 1924.
6. McCurdy, H. G. Group and individual variability on the Goodenough Draw-A-Man Test. *Educ. Psychol.*, 1947, *38*, 428-436.
7. Pikunas, J. *Fundamental child psychology.* Milwaukee: Bruce, 1957.
8. Pikunas, J. *The graphoscopic scale: a multi-dimensional projective technique.* Detroit: University of Detroit Press, 1959.
9. Pikunas, J. and Albrecht, E. J. *Psychology of human development.* New York: McGraw-Hill, 1961.

STRUCTURED PERSONALITY TESTING IN THE AGED: AN MMPI STUDY OF THE GERONTIC POPULATION[1]

WENDELL M. SWENSON

Mayo Clinic, Rochester, Minnesota

Problem

Careful perusal of the literature reveals no personality studies among the aged in which the Minnesota Multiphasic Personality Inventory (MMPI) was used as a primary psychometric device. Calden and Hokanson[1] included 15 male tuberculous patients aged 60 to 69 in a study measuring the influence of age on MMPI results. The purpose of the present study is twofold: to indicate how well elderly individuals can be tested with presently available structured personality tests, and to describe a sample of the nonhospitalized aged population in terms of the Minnesota Multiphasic Personality Inventory.

Procedure

The MMPI data described here were obtained as a part of a larger psychometric study of a gerontic population. Two hundred ten individuals aged 60 years and more cooperated in completing a brief attitude questionnaire. These subjects were asked also to complete the group form of the MMPI, but very little individual urging was done. From the 210, 95 (45.1 per cent) completed the MMPI and obtained valid profiles defined as one in which the "cannot-say" score did not exceed 110 and the raw F score was not greater than 21. Of the 115 subjects in the study not completing the MMPI, 70 refused to take the test, 39 began the test but stopped before completing it, and 6 completed the test but with "cannot-say" scores above the upper limit for validity. As a group, subjects who cooperated in taking the MMPI were younger than those who did not (median ages 71.4 and 74.8 years); and they were more often gainfully employed rather than confined to a rest home.

The 95 tested subjects came from three separate sources: (1) so-called "Golden Age Clubs," (2) homes for the aged, and (3) their several places of employment. The men numbered 31 and the women 64. No attempt was made to structure the sample so as to include or eliminate any particular type of classification of the aged, except that those with known mental disorders intense enough to warrant a psychiatric diagnosis were eliminated. All the subjects in the "Homes for the Aged" group were confined in the home primarily for physical incapacity.

The examiner met with the subjects individually or in small groups at their clubs or resident homes or elsewhere, and described the project as a "study of interests and attitudes of the aged." The group form of the MMPI was used exclusively in the testing.

Results

The coding system used in this investigation is that described by Hathaway and Meehl[4]. Each of the 10 scales is represented at least once as high-point in the codes of both sexes, with the exception of scale 7, which has no high point in the male sample. Scales 0[2], 1, 2, and 3 are the most frequently occurring high point scales in both sexes. The 0 scale appears more frequently by proportion in female subjects than in the male group. Combinations of scales 4 and 9 in either sex as high points are conspicuous by their absence. These two scales (4 and 9) appear with relatively high frequency among the sample[3] of 258 male and 368 female adult normals in Minnesota. Comparison of this group with the normative group suggests greater statistical abnormality in the gerontic group (more codes with scales > 70).

In the male gerontic sample, eight (25.8 per cent) of the 31 profiles have at least one score equal to or greater than T-score 70, and therefore are identified as "prime"

---

[1]Read at the meeting of the American Psychological Association, Cincinnati, Ohio, September 4, 1959.

[2]0 refers to the Social Introversion Scale.

codes. In 6 (75 per cent) of these "prime" codes, the 2 scale is one of the scores equal to or greater than 70. In the female gerontic group are 20 prime codes which constitute 31.9 per cent of the entire female sample. These prime codes appear in all 10 of the clinical scales except the 4 scale, and show no particular point of concentration.

The relative frequencies of the various high-point codes in the male and female samples are presented in table 1. Four frequently occurring one-point codes are

TABLE 1. FREQUENTLY OCCURRING ONE POINT MMPI CODES FOR 31 MALE AND 64 FEMALE GERONTIC NORMALS

High-Point Code					
Scale	Males Cases	%	Scale	Females Cases	%
3	8	25.8	0	17	26.5
2	7	22.6	1	11	17.2
1	4	12.9	2	9	14.1
0	4	12.9	3	6	9.3
			6	6	9.3

identical in the male and female groups, but are reversed in rank (male: 3, 2, 1, 0; female: 0, 1, 2, 3). No two-point codes appear with sufficient frequency to be identified as typical of either the male or female group. The 9 scale is the most frequently occurring low point, and is found in 10 (32.1 per cent) of the male subjects and in 23 (35.9 per cent) of the female group. The next most frequent low points in the profiles are the 6 scale and the 4 scale.

Hathaway and Monachasi[4], in their studies of 4000 midwestern ninth graders, concluded that the median adolescent profiles tended to be in the direction of "behavior disorders" or of psychotic trends rather than toward neurotic patterns. Definitely, the reverse was true in this gerontic group. The median profiles here were commonly of a neurotic pattern, and only rarely gave evidence of psychosis or behavior disorders. The most typical combination of coded high and low points for both sexes in the adolescent group (Hathaway and Monachasi) was 9-2. In the gerontic group that combination is 3-9 for the males and 0-9 for the females.

It is of interest to compare this sample with one other group on which there are MMPI data available. Schiele and Brozek[5], analyzing the MMPI profiles of 32 adult males subsequent to a 6-month period of semistarvation, found a "mean profile" strikingly similar to the median profile obtained in the present investigation. The high points of the semistarvation group were scales 2, 3, and 1, and the low point of the typical code was scale 9. The subjects in the semistarvation group had undergone physiologic stress which, among other things, probably had threatened their security. Granick[2], in describing the psychopathology of aging, stated that "the outstanding factor in personality breakdown in old age seems to be insecurity." The possible psychologic similarity between these two groups at present can only be conjectured.

In 1956, 100 female geriatric patients in a state mental hospital were evaluated by the writer through the use of structured psychometric devices. Sixty-six of these patients cooperated in taking the MMPI and produced valid profiles. Table 2 shows the means and standard deviations of this group on each of the MMPI scales in comparison with means and standard deviations of the 95 subjects of the present study. Because of apprehension concerning the use and comparison of "composite" MMPI profiles, no statistical comparison of these two groups is made here. Certain trends, however, are evident: (1) The state hospital group appears to be more variable in its personality characteristics than the so-called normal gerontic group. (2) The T-scores of the clinical scales in the MMPI tend to be higher among the state hospital geriatric group. (3) The elevations tend to be higher in the psychotic end of the profile (scales 6 and 8, particularly) in the state hospital group.

TABLE 2. TABLE OF MEANS AND SIGMAS OF MMPI T-SCORES OBTAINED FROM 66 FEMALE STATE HOSPITAL PATIENTS AND 95 NONHOSPITALIZED AGED

Scale	Means		Standard Deviation	
	Geriatric patients	Gerontic normals	Geriatric patients	Gerontic normals
?	7.77	11.03	20.82	23.05
L	7.50	6.137	3.31	3.59
F	9.86	4.432	7.85	3.11
K	15.91	14.389	5.87	4.65
1 (Hs)	62.23	57.95	12.78	12.768
2 (D)	65.18	59.21	13.38	11.74
3 (Hy)	61.26	55.91	13.19	11.69
4 (Pd)	62.71	51.02	13.07	9.137
5 (Mf)	57.39	55.62	12.67	10.49
6 (Pa)	65.85	52.43	13.58	8.74
7 (Pt)	59.00	53.17	11.49	10.41
8 (Sc)	65.70	52.45	14.80	9.73
9 (Ma)	57.24	47.22	11.89	10.04
10 (Si)	60.07	56.54	8.63	7.72

SUMMARY

Without much individual pressure 95 of 210 subjects more than 60 years old were persuaded to take the MMPI group test. The typical or median profile for this group is a neurotic one with an absence of evidence of psychotic or behavior disorder tendencies—a finding in reverse of that demonstrated in a large group of adolescents, but strikingly similar to a small group of males subjected to a period of semistarvation.

A distinct sex difference in the group occurs in the frequency of the appearance of the 0 scale as the high point in the profile, its presence being much more common among females. This finding is similar to that from the ninth-grade students studied by Hathaway and Monachasi [4].

Scales 4 (Pd) and 9 (Ma) or any combination of them occur rarely as high-point. These two scales stand out clearly as the most common low-point codes of both sexes. Both sexes have fewer codes with no high point than are found in samples of younger normal males and females.

Approximately 25 per cent of the males and 30 per cent of the females have prime coded profiles indicating T-scores of 70 or greater on at least one scale. The 2 (D) scale tends to appear more frequently as a codable high point among males than among females.

REFERENCES

1. CALDEN, G. and HOKANSON, J. E. The influence of age on MMPI responses. *J. clin. Psychol.*, 1959, *15*, 194-195.
2. GRANICK, S. Studies in the psychology of senility: A survey. *J. Gerontol.*, 1950, *5*, 44-58.
3. HATHAWAY, S. R. and MEEHL, P. E. *An Atlas for the clinical use of the MMPI.* Minneapolis: The University of Minnesota Press, 1951, p. xxiv.
4. HATHAWAY, S. R. and MONACHASI, E. D. (Eds.). *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: The University of Minnesota Press, 1953, p. 96.
5. SCHIELE, B. C. and BROZEK, J. "Experimental neurosis" resulting from semi-starvation in man. In Welsh, G. S., & Dahlstrom, W. G. (Eds.). *Basic readings on the MMPI in psychology and medicine.* Minneapolis: The University of Minnesota Press, 1956, pp. 461-483.

RORSCHACH RESPONSES OF CHILDREN EXHIBITING PSYCHOGENIC AUDITORY SYMPTOMS

FRANK KODMAN JR. AND JERRY E. WATERS

University of Kentucky

Problem

An interesting problem observed by the clinician is the youngster, preschool or school age, who reacts to audible sound as though he were hard of hearing when in fact his hearing is normal or near normal. Although this youngster will frequently show a discrepancy of 15 db or more between various threshold measurements, it is not sufficient simply to label him as a *functional* hearing loss case. We would argue on clinical grounds that functional hearing loss, when properly evaluated, is a manifestation of psychopathology.

Juers[2], Kodman and Pattie[3], Kodman, *et al*[4], Kodman and Blanton[5], and Myklebust[6] have called attention to psychogenic deafness in children where the cardinal auditory symptom was usually a subnormal response to conversation with a discrepancy between the pure tone thresholds and the speech thresholds. Psychogenic deafness in children as a manifestation of psychopathology warrants further inquiry into the personality dynamics involved *in addition* to the measurement of the degree of functional hearing loss by auditory tests. This study investigates some of the personality characteristics of early psychogenic deafness by means of Rorschach analysis.

Procedure

Eight children, four males and four females, were diagnosed as psychogenic largely on the basis of their general behavior and a 30 db or greater loss between conventional pure tone audiometry and conventional speech audiometry. They were matched for age and sex with eight hard-of-hearing children who showed an average 30 db binaural loss with a range from 20-50 db. The psychogenic group showed essentially normal hearing (loss no greater than 5 db) by speech audiometry. The average functional loss for the Psychogenic group was 30 db with a range from 20-55 db.

The children were screened for mental retardation and aphasia by means of an extensive case history, parent interviews, educational and audiological evaluation, and psychological tests including the Rorschach, Columbia Mental Maturity Scale, Bender Gestalt and Draw-A-Man tests. The mean age of each group was 11 years; range 8-14 years. Each group fell within the normal IQ range with a mean of 94.0 SD $\pm$ 6. The children were selected carefully from our diagnostic caseload over a three year period. Only the more meaningful characteristics of the data are reported here.

Results

The average number of Rorschach responses for the Control and Psychogenic groups was essentially the same. The mean R values of 23.25 and 23.38 compare favorably with the mean R of 27.15 shown for normal children by Thetford, Molish and Beck[8]. The mean numbers of W, D, and Dd, frequently called the approach, were 3.88, 16.88, and 2.50 for the Control group as compared with 2.13, 19.38 and 1.88 for the Psychogenic group. The Wilcoxon Matched-Pairs Signed-Ranks Test[7] showed no significant differences between these factors.

The Control group exhibited a mean Z value of 14.13 which approximates the Thetford, Molish and Beck Z of 11.05 and is higher than the mean 4.06 Z value obtained for the Psychogenic group. A total ranks value (T) of 4.0 indicates that the difference between the Z values for the two groups is significant at the .05 level.

Constriction of the responses of the Psychogenic group as indicated by the low Z value, is further corroborated by few human movement and color responses. Three

psychogenic subjects gave color determinants. When color was reported, it was controlled, that is, all color responses were FC for the Psychogenic group. In contrast, four of the hard-of-hearing control subjects utilized color with a tendency to use uncontrolled C, CF+ and CF− types of determinants. The Ranks Test was not applied since too few subjects produced color responses. However, the mean sum C value of 2.06 of the Control group compares favorably with the sum C norm of 2.02 reported by Thetford, Molish and Beck for their total group of normal children. Both groups made infrequent though normal use of white spaces.

Shading or Y responses were minimal for the Psychogenic group. Only one shading response (YV) was produced. Shading responses were utilized more extensively in the Control group with five subjects producing either a Y, YF or FY response. However, only the FY+ response was given by two or more of the hard-of-hearing control subjects. Again, the most outstanding feature is the constriction shown by the Psychogenic group. In regard to the M/C ratio, two of the control subjects and three of the psychogenic subjects failed to produce either an M or C.

The mean F+ percentages of 55.88 and 74.50 as well as the mean extended F+ percentages of 59.75 and 76.20 are much higher for the Psychogenic group. The differences in both percentages between the two groups are significant at the .02 level according to the Ranks Test. The mean F+ percent for the Psychogenic group more closely approximates a mean F+ percent of 70.05 with a standard deviation of 11.54 reported by Thetford, Molish and Beck. Therefore, it seems that the Psychogenic group is demonstrating good form perception indicative of good ego strength and respect for reality. In contrast, the mean F+ percent of the Control group is beyond one standard deviation of the expected mean. For adult norms, Beck[1] uses the third standard deviation as the critical F+ percent for the healthy personality. The Control group F+ percent of 55.88 falls within the second standard deviation and may be interpreted as being in the direction of the neurotic personality.

The constriction of the responses of the Psychogenic group is again evident in the high F percent of 94.75. Thetford, Molish and Beck found a mean of 79.80 for F percent which closely parallels the mean of 79.00 obtained for the Control group. A total ranks value of 1 (P = .05) indicated that the F percent obtained for the Psychogenic group was significantly greater than the F percent obtained for the Control group.

Summary

This investigation was concerned with the personality characteristics of a sample of children diagnosed as psychogenically hard of hearing. They were compared with a Control group of hard-of-hearing children. The Psychogenic group manifested certain hysteroid type characteristics. An extended report may be obtained from the authors.

References

1. Beck, S. J. *Rorschach's Test: II. A Variety of Personality Pictures.* New York: Grune and Stratton, 1949.
2. Juers, A. L. Puretone threshold and hearing for speech—diagnostic significance of inconsistencies. *Laryngoscope*, 1956, *66*, 402-409.
3. Kodman, F. and Pattie, F. A. Hypnotherapy of psychogenic hearing loss in children. *Am. J. clin. Hypnosis*, 1958, *1*, 9-13.
4. Kodman, F., Lieberman, L., Byers, V. W. and Farquharson, C. Some implications of hearing defective juvenile delinquents. *J. except. Child.*, 1958, *25*, 54-56, 67.
5. Kodman, F. and Blanton, R. L. Perceptual defense mechanisms and psychogenic deafness in children. *Percept. Mot. Skills*, 1960, *10*, 211-214.
6. Myklebust, H. R. *Auditory Disorders in Children.* New York: Grune and Stratton, 1954.
7. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw Hill, 1956.
8. Thetford, W. N., Molish, H. B. and Beck, S. J. Developmental aspects of personality structure in normal children. *J. Proj. Tech.*, 1951, *15*, 58-78.

NUDITY IN HUMAN FIGURE DRAWINGS

GARY M. FISHER[1]

Fairview State Hospital, Costa Mesa, California

PROBLEM AND PROCEDURE

Machover [2, p. 76] has suggested that those who underclothe the drawing of the human figure (designated as body narcissists) have "more than the average degree of sexual maladjustment", and also that sexual maladjustment is indicated when the opposite-sex figure is drawn first. This study relates the sex of the first-drawn figure to the presence of nudity in the figure drawings of 1154 white male adolescents (14 to 19 years) who were incarcerated for delinquent behavior. The great majority of Ss had histories of chronic delinquency, but no cases involving sexual crimes or deviations were included in the study.

Each drawing was placed in one of the following five categories: fully clothed; partially nude (bathing suits, trunks, bare chest, etc.); nude with genitals; nude without genitals; outline of figure only (neither clothed nor nude). A test of the significance of the difference between independent proportions was performed for each category between Ss who drew the male figure first and those who drew the female figure first.

RESULTS

The male figure was drawn first by 85.3% of the Ss, a finding which is in agreement with the results of previous studies. [1] Table 1 presents a comparison of the drawings of Ss who drew the male figure first and those who drew the female figure.

TABLE 1. DIFFERENCES IN INCIDENCE OF NUDITY IN FIGURE DRAWINGS ACCORDING TO THE SEX OF THE FIRST-DRAWN FIGURE

Drawings	Male Figure Drawn First (N 984) N	%	Female Figure Drawn First (N 170) N	%	Critical Ratio	*P*
Fully clothed						
Male figure	843	85.7	127	74.7	3.62	.001
Female figure	829	84.3	113	66.5	5.17	.001
Partially nude						
Male figure	56	5.7	17	10.0	2.04	.05
Female figure	53	5.4	26	15.3	4.66	.001
Nude with genitals						
Male figure	6	.6	8	4.7	4.50	.001
Female figure	26	2.6	13	7.7	3.40	.001
Nude without genitals						
Male figure	32	3.3	13	7.7	2.75	.005
Female figure	36	3.7	13	7.7	2.45	.01
Outline of figure only						
Male figure	47	4.8	5	2.9	1.12	N.S.
Female figure	40	4.1	5	2.9	.75	N.S.
Ss indicating nudity in one or both drawings	148	15.0	62	36.5	6.72	.001

first. Ss who drew the female figure first indicated nudity significantly more frequently in both the male and female figures in all three categories of nudity (partially nude, nude with genitals, nude without genitals) than did Ss who drew the male figure first. Of the Ss who drew the female figure first, 36.5% gave some indication of

[1]Appreciation is expressed to Connell Cowan and John Stamford for their assistance in this study.

nudity in one or both of their drawings whereas only 15% of the Ss who drew the male figure first did so. In addition, clothed figures were drawn significantly more frequently by Ss who drew the male figure first than by those who drew the female figure first, and no differences existed between the two groups with respect to the frequency with which outline-figures (neither clothed nor nude) were drawn. Thus, two drawing characteristics which purportedly reflect sexual maladjustment have been shown to be positively related to each other.

Summary

Indications of nudity in the human figure drawings of 1154 white male adolescent delinquents were related to the sex of the first-drawn figure. Ss who drew the female figure first had a significantly greater incidence of nudity (partially nude, nude with and without genitals) in both the male and female figures then did Ss who drew the male figure first. Two figure drawing characteristics purportedly reflecting sexual maladjustment have been shown to be positively related to each other.

References

1. Brown, D. G. and Tolor, A. Human figure drawings as indicators of sexual identification and inversion. *Percept. Mot. Skills*, 1957, 199-211, (Monogr. Suppl. 3).
2. Machover, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Thomas, 1949.

PERFORMANCE CHANGES IN REPEATED JUDGMENTS OF SCHIZOPHRENIC VERBAL RESPONSES[1]

STANLEY BLUMBERG

University of Texas, Southwestern Medical School

Problem

This paper extends the study of judgments of disorganization in schizophrenic thinking into the area of learning by investigating the effects of knowledge of results and of practice in the absence of external reinforcement on clinical judgments. Since there is some evidence from studies[2] in the area of perceptual learning that absolute judgments by the method of single stimuli improve with practice even without correction, or knowledge of results, clinical judgments might also be expected to become more accurate with practice. Also, since all experiments providing a contrast between practice with and without knowledge of results show a clear superiority of the condition with knowledge, clinical judgments should improve faster under the latter condition. A review of a wide variety of judgmental and perceptual-motor studies[1] using knowledge of results for reinforcement presents the general conclusion that learning is faster and reaches a higher level with knowledge of results, with more specific knowledge leading to more rapid improvement than less specific knowledge.

This study was designed to test the effects of three experimental conditions on repeated clinical judgments. The hypotheses were as follows: (a) Practice in making judgments, in the absence of information as to their accuracy, would nevertheless

[1]This study is part of a larger project being conducted at Northwestern University under the direction of Professor William A. Hunt through Contract 7onr450(11) with the Office of Naval Research. The opinions expressed are those of the author and do not represent opinions or policy of the Naval service. An extended version of the study was submitted to the Graduate School of Northwestern University in partial fulfillment of the requirements for the Ph.D. degree. The author gratefully acknowledges the guidance and assistance of Dr. W. A. Hunt, committee chairman, and Drs. R. I. Watson, Janet A. Taylor, and W. F. Hill.

lead to more rapid, reliable, and valid judgments, thus indicating that learning had taken place. (b) Specific knowledge of the correct judgmental answers would result in even more rapid, reliable, and valid judgments. (c) General, or less specfic, information as to the correct answers would lead to an intermediate position, with respect to learning correct judgmental responses, between practice without feedback information and the specific reinforcement condition. (d) These three conditions would result in transfer of training (more rapid, reliable, and valid judgments) when new stimuli were presented to be judged.

Method

Subjects. Three groups of judges, 30 in each group, comprised the subjects (*Ss*) in the study. All *Ss* were drawn from one introductory psychology class for college undergraduates and were assigned randomly to the groups. All *Ss* were naive, having never previously served in a judgment experiment.

Procedure. Each *S*, tested individually, rated on a 7-point scale for severity of disorganized thinking 21 schizophrenic verbal responses to vocabulary test items previously standardized by Hunt and Arnhoff [3]. These schizophrenic responses, selected to cover the entire stimulus range with three stimuli roughly at each of the seven scale points, were presented one at a time by a slide projector to which a timer was connected. The apparatus is described elsewhere [5]. The series of 21 stimuli was repeated six times in a different random order for each presentation. Four practice stimuli not contained in the regular series preceded the experiment proper.

Group I *Ss* received only practice in making their judgments; i.e. they were given no information as to the accuracy of their ratings. Group II *Ss* were told after each rating whether their judgment was correct, too high, or too low. The third group was given the correct scale value for each item after each judgment was made. After the six training trials all *Ss* were given a new list of 21 items, equated in scale values to the first list, which they rated twice without receiving any knowledge as to the accuracy of their judgments.

Measures. Three measures of performance on the first six trials were obtained as indices of learning. These were latency, or time of judgmental response, reliability of the ratings as measured by interjudge agreement, and accuracy, or validity, of the ratings as measured by product moment correlations with the established standardized values. These same measures were used for evaluating transfer of training.

Results and Discussion

As indicated in Table 1, mean judgment times over the first six trials dropped essentially uniformly for all three groups of judges suggesting that learning occurred

Table 1. Mean Judgment Times in Seconds, Mean Reliability Coefficients and Mean Validity Coefficients Over Trials

Group	Trials 1	2	3	4	5	6	7	8
I	8.0	5.3	4.2	3.7	3.5	3.3	8.0	4.5
II	7.9	5.4	4.8	4.0	3.7	3.4	7.6	5.0
III	7.5	5.2	4.4	4.0	3.7	3.3	7.5	5.0
Mean Reliability Coefficients								
I	.808	.805	.831	.817	.815	.816	.799	.812
II	.741	.791	.802	.867	.901	.903	.798	.818
III	.768	.852	.874	.923	.931	.946	.824	.822
Mean Validity Coefficients								
I	.776	.769	.800	.787	.788	.782	.749	.766
II	.697	.781	.830	.866	.902	.903	.751	.780
III	.730	.852	.872	.919	.930	.946	.800	.797

at the same rate for all three groups. Analysis of variance of the individual mean judgment time scores revealed an absence of differences between groups, with *trials* as the only significant variable affecting speed of judgment. Since the time scores on trial seven are the same as on trial one, it should be concluded that the three conditions used in the original training situation had no differential effects on the amount of time taken to judge new stimuli.

Reliability of the ratings, measured by interjudge agreement, increased over the six training trials, as indicated in Table 1. As predicted, Group III achieved greater agreement by the sixth trial than did Group II, which in turn showed more agreement than did Group I. Analysis of variance of the Z transformations of the individual reliability coefficients indicated significant differences between groups and trials, and interaction between trials and groups. This suggests that specific knowledge of results leads to higher reliability of judgments than does less specific knowledge which in turn results in higher reliability than does no information (practice alone).

In spite of some increase in interjudge agreement over trials for Group I the difference between their reliability on trial three (.831) and on trial one (.808) is not statistically significant ($t = 1.15$). This leads to the conclusion that in the absence of some kind of information as to the accuracy of their judgments, *Ss*' judgments do not become significantly more reliable with practice at judging the same stimuli. A conclusion that these *Ss* are not learning *anything*, however, is not warranted by these results. What is meant by learning must be specified, and the most likely explanation relating the time score data to the reliability data is that all three groups did, in fact, learn but that Group I learned something different from what Groups II and III learned. Groups II and III learned the standardized values for the stimuli. *Ss* in Group I learned their own responses to the specific stimuli presented over repeated trials, which accounts for their more rapid response times over trials. This result replicates that found in the Hunt and Blumberg[5] study of repeated judgments.

A comparison of the mean reliability coefficients on trial seven with those on trial one reveals improved performance (positive transfer of training) for both reinforced groups but not for Group I. Group II improved in reliability from .741 to .798 ($t = 1.97$, $p < .05$), while Group III went from .768 to .824 ($t = 3.06$, $p < .005$). This result suggests that, in the absence of information as to the accuracy of their judgments, practice in making clinical judgments of this kind does not lead to improvement (higher reliability) in judging new but similar material. On the other hand, such improvement does appear to occur following practice with such information, with the greatest improvement resulting from the *specific* knowledge of results training condition.

The results with respect to validity, the correlation of each *S*'s judgments with the standardized values of the stimuli, resemble those obtained using reliability as the measure of performance. Group III showed the most rapid learning, or improvement in terms of the validity of their ratings, over the six training trials, followed by Group II, followed by Group I as shown in Table 1. Analysis of variance showed significant differences between groups and trials. Again, there was no significant improvement in Group I's ratings.

Apparently, while each *S* in Group I is learning his own responses to the specific stimuli, and thus becoming more consistent with himself, he is not becoming more consistent with his own group of judges (reliability in this study) nor with the group of experienced clinicians upon whom the items were standardized (validity in this study). This is in accord with the finding of Seashore and Bavelas, reported in Ammons'[1] review, that repeated practice in the absence of any knowledge of results may result in an increasing self-consistency although no improvement in performance in terms of accuracy with the experimenter's standard occurs. Thus the hypothesis, based on some of Gibson's[2] findings, that Group I *Ss* would improve in their reliability and validity of judgments in the absence of knowledge of results was not

confirmed. The hypothesis, or assumption, was that during the first few trials *Ss* would learn more about the range of stimuli and correct prior judgments which they felt to be in error as the context became better established.

As for transfer of training, using mean validity coefficients as a measure of performance, we find Group II obtaining .697 on trial one and .751 on trial seven, a difference which is not significant at the .05 level ($t = 1.57$, $p < .10$). Group III, however, improved from .730 to .800, a difference which is significant ($t = 3.24$, $p < .005$). Again, Group I showed no improvement, and in no case was performance on trial eight superior to that on trial two. Apparently, the specific reinforcement led to the greatest amount of learning (learning the "correct" scale values for the first list of items) and also resulted in the highest performance, as measured by both reliability and validity, when *Ss* were presented with the new list of stimuli on trial seven.

In conclusion, the results support the hypothesis that practice at rating "clinical stimuli" of this type leads to more rapid judgments upon repeated presentations, but does not lead to more reliable or valid judgments of either the same or new stimuli. On the other hand, practice with a *general* knowledge of the relative accuracy of one's judgment results in more reliable and valid ratings of the same material, but this may not transfer to new stimuli. Practice with *specific* knowledge of the correct rating apparently results in more reliable and valid judgments of not only the same stimuli, but also highly similar new stimuli.

Summary

Three groups of naive judges, 30 to a group, rated 21 schizophrenic vocabulary test responses on a 7-point scale for disorganization of thinking. The judgments were repeated over six trials. One group received only practice in making their judgments; *i.e.*, they were given no information about the accuracy of their ratings. The second group of subjects were told after each rating whether their judgment was correct, too high, or too low. The third group was given the correct scale value for each item after each judgment was made. After the six training trials all subjects were given a new list of 21 items, equated in scale values to the first list, which they rated twice without receiving any knowledge as to the accuracy of their judgments.

The results indicated that practice, without information as to the accuracy of judgments, did not lead to improvement, *i.e.*, higher reliability and validity, in clinical judgments; but that both specific and less specific knowledge of the correct ratings did. Only the training condition involving specific knowledge resulted in transfer of training, *i.e.*, statistically significant improvement in both reliability and validity of judgments of new material following original training.

References

1. Ammons, R. B. Effects of knowledge of performance: a survey and tentative theoretical formulation. *J. gen. Psychol.*, 1956, *54*, 279-299.
2. Gibson, Eleanor J. Improvement in perceptual judgments as a function of controlled practice or training. *Psychol. Bull.*, 1953, *50*, 401-431.
3. Hunt, W. A. and Arnhoff, F. N. Some standardized scales for disorganization in schizophrenic thinking. *J. consult. Psychol.*, 1955, *19*, 171-174.
4. Hunt, W. A. and Arnhoff, F. N. The repeat reliability of clinical judgments of test responses. *J. clin. Psychol.*, 1956, *12*, 289-290.
5. Hunt, W. A. and Blumberg, S. Manifest anxiety and clinical judgment. *J. clin. Psychol.*, 1961, *17*, 8-11.

BELOW CHANCE PERFORMANCE ON A VERBAL TEST USING PAIRED CONCEPTS

E. FRANCESCO

Box 601, Ithaca, N. Y.

INTRODUCTION

In the design of tests that have dichotomous response alternatives, the usual procedure is to balance the test so that it has as many items with one alternative correct as the other. That is, in a test with true-false type answers, there are generally as many items with correct true as correct false answers. The logic underlying this design is that when the respondent is not answering on the basis of his knowledge, he should be getting no more than half the answers correct on the basis of guessing, or by "chance." The design is assumed to be valid for any respondent strategy of answering if the items are presented in random order. Guessing all true, all false, different or equal proportions of true or false, at random or by any system, should not improve the score of the respondent. There is an additional assumption, however, that must actually be satisfied for this to be true, namely, that the test also should be balanced with regard to the difficulties of the items associated with the two alternatives.

It is well to illustrate a strategy that would beat the system if the alternatives are not balanced with regard to difficulty. Suppose that in a hypothetical test the items that have correct answers that are true are easier than those that are false. A respondent might take the test answering all the items he is quite sure about *first*. Suppose that he finds that of 100 items he has answered 25 true and 5 false. Then, if he assumed that there is a balanced number of true and false items in the test, he can simply answer all remaining items as false and get a better score than by guessing each individually. By the former strategy he would get a score of 75 correct, by the latter only 65 correct. Of course, it is the choice of strategy itself that involves uncertainty for the respondent at this point. It might be easier to reason that the test is biased in the direction of inclusion of more true items than false ones, since the number of true ones the respondent recognized far exceeded the number of false ones. With this strategy, i.e., by answering all remaining items as true, the respondent would get a score of 50, which means that his strategy actually counteracted any advantage that he would have achieved on the basis of his actual knowledge. The "safe" strategy is the one of guessing items individually, which presumably in the absence of knowledge is a random set of responses.

The parallel situation occurs when tests are unbalanced in the number of items, assuming the difficulties of the items associated with the two alternatives are equivalent. In this case, the best strategy for a test, once the proportion of true and false answers is assessed, would be to assign all remaining answers to the response alternative with the larger proportion of known correct answers.

ILLUSTRATIVE MATERIALS

In the development of the Verbal Power Test of Concept Equivalents (*VPT*), these considerations of difficulty entered the design prominently. The objective of balance in number and in difficulty did not appear possible on the basis of actual experience. The instructions for the *VPT* define the task as the recognition of concept pairs that are the *Same* (e.g., neat-tidy) or *Different* (e.g., foot-shoe). As has been previously reported [2], the task of recognition of *Different* pairs was judged easier than that of recognition of *Same* pairs on the empirical basis of respondent performance. Thus, the *VPT* forms were designed to each have 63 *Same* items and 42 *Different* items; the logic simply was to provide for a balanced appearance if one is answering many of the *Different* items correctly and has insufficient knowledge to answer the *Same* items and must resort to guessing or to choosing a test strategy

for answering. The design of the test maintained the same ratio of *Same* and *Different* in a revision, but the ratio was increased to 46 to 29 when the number of items for each of the forms was reduced in accord with experience on the test, and both forms of the test were placed on a single IBM response form.[1] The experience discussed below refers to the 75 items that compose the A form of the test.

Interest in the question of below chance performance arose out of a sample of respondents in which a substantial proportion of *VPT*s were completed at or below the manifest chance level, when this was assumed to be half of the dichotomous responses answered correctly.[2] In this sample about 40 per cent of respondents (Table 1) had 39 or fewer items correct out of the possible 75. About 10 per cent had 34 or fewer correct. The question that came to the fore was the obvious one. Since the *VPT* forms had been devised on "average" subjects, and this group reasonably fits the description[3], what were the characteristics of responses at or below the manifest chance level?

TABLE 1. MEAN *Same* AND *Different* CORRECT RESPONSES BY TOTAL NUMBER CORRECT

Total number correct	Number of cases	Mean *Same* correct	Mean *Different* correct
25-29	3	9.0	17.3
30-34	20	15.3	17.2
35-39	54	20.5	17.3
40-44	46	22.8	19.1
45-49	29	27.2	19.9
50-54	19	28.5	22.7
55-59	14	32.3	24.7
60-64	6	34.5	26.7
65-69	3	38.0	28.0

The data from Table 1 are plotted in Figure 1. It is seen that while the mean number of *Different* items that would be expected to be answered correct at the chance level is 14.5, the mean number correct does not go below 17.2. The mean number of *Same*-items that would be expected to be answered correct would be 23, but the mean number actually gotten correct goes well below this, and in fact persons with from 40-44 correct in the total score are performing with a mean number *Same* correct just slightly under 23. Because of the apparent fixed minimum number of *Different* items correct, for persons operating below the manifest chance level on the total scores the ratio of correct *Same*-to-*Different* is inverted relative to the ratio of available items.

DISCUSSION

We may now look to some of the possible interpretations of the findings in terms of the strategies of response.[4] What does the task of recognizing *Same* or

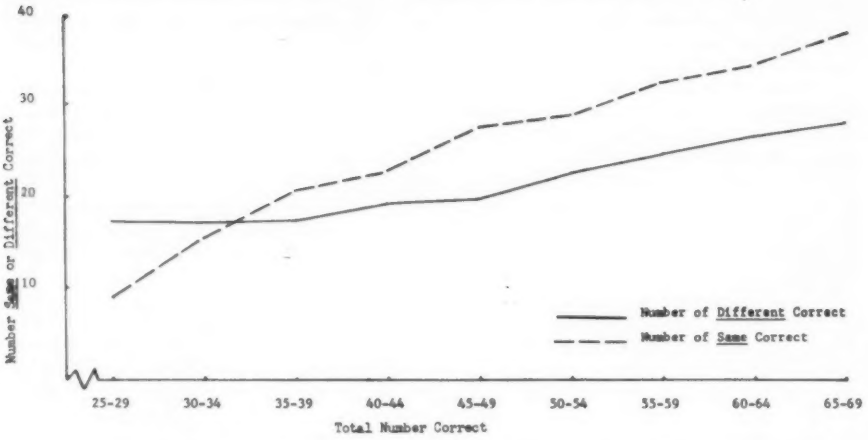
[1] Length of the test was set primarily on considerations of improvement of correlation to external measures. In the current length of forms (75 items each) improvement is not marked by doubling the length of the test. For example, in a sample of high performing housewives in the Broome County Extension Service leader training program (N = 143), the correlations of forms A and B to the *Quick Word Test* were .77 and .80 respectively, while the combined forms correlated at .83. The between forms correlation was .80 for this restricted variance group. A supply of the new IBM forms of the *VPT* are available for research in studies of adult intelligence by writing the author (Box 601, Ithaca, N. Y.).

[2] These data were provided through the generous cooperation of Dr. James H. Panton, Supervisor, Reception Center, State of North Carolina Prison Department.

[3] This group had a mean total IQ on the WAIS of 99.0, but the SD was very small (10.4); the population SD for the WAIS is given as 15, indicating the highly restricted variance in this sample.

[4] Many of the comments on strategy arise, of course, from interviewing respondents at various times. There are limitations to this kind of procedure, however, best described as the intrusion of the respondent's concern with the interpretation of the concepts, or "why I got that one wrong."

Figure 1. Mean *Same* and *Different* Correct Responses by Total Number Correct



Different concepts actually involve? First, there are two clear conditions of accurate response when both concepts are known. Second, there is the possibility that only one is known or both are not known, and then it would be expected that some strategy of response would operate. If neither is known, there is no basis for judgment, and thus any guess might be appropriate. If one is known but not the other, this might well provide the situation for an assumption that it is a mismatch, since the compared concept doesn't mean to the respondent what the known concept means. The evidence, however, seems to point simply to a tendency to answer *Different* when one does not know that they are the same. Thus, decrease in total scores on the *VPT* appears to be consistently accompanied by an increase of incorrect or *Different* responses on the *Same* items. If scores even lower than those found were to occur, extending the observed trend we would expect them to occur because more *Same* items were being responded to as *Different*. Further, if the trend were correctly assessed, the tendency should be for more *Different* items to get correctly answered as an artifact of the increased lack of knowledge. In Figure 1 an increase in the number of *Different* items answered correctly is not noticeable, but it should be recognized that, for example, in the group having 30-34 total number correct a mean score of 15.3 *Same* items correct means that 30.7 are incorrect, or about 67 per cent of the *Same* items are answered as *Different*. Only 17.2 or 59 percent of the *Different* items are answered as *Different*. In other words, at this level respondents are answering *Different* items as *Different* less frequently than they do the *Same* items, and although only three cases are involved, this is even more accentuated for the group having 25-29 total number correct.

One other way to examine the concomitants of lack-of-knowledge as a variable is to select the respondents on the basis of another criterion. Here we may look at the low end of the distribution on the WAIS, and we find that for the nine cases with full scale scores under 80, the total mean number correct was 32.5, and these were divided into 12.1 *Same* and 20.4 *Different* correct. About 74 per cent of the *Same* items are being answered as *Different* while only about 70 percent of the *Different* items are being answered in this manner. The *Different* items have, so to speak, a resistance to being seen as *Different* that is greater than that of the *Same* items. Presumably, however, if persons were chosen who had insufficient knowledge, *Same* and *Different* items would be responded to identically.

Here, however, the essential point is clear. The test is operating with interpretable characteristics below its manifest chance level. Some limitations need to be

pointed out. A person who operates on a random basis will get a better score (mean 37.5) than persons who have some knowledge and operate on the kinds of bases indicated, namely to assume generally that when only one concept is known or both are not known, they are *Different*. On the other hand, dichotomizing persons getting scores of 37 and 38 (manifest chance level) by the ratio of *Same* to *Different* correct responses did not indicate any corresponding difference on WAIS scores.

It should be noted that the proportion of *Different* scores seems to have an upper limit, possibly at about 75 per cent. (No case of all *Different* responses has been encountered yet.) It may appear unreasonable to respondents as a general strategy of test taking to believe that all the items would have responses of *Different*, and thus a counter strategy may require a certain amount of guessing that items should be scored as *Same*. According to informal interviews, this seems to be a secondary consideration, ostensibly placing a limit on the tendency to respond to items individually as *Different*. This same kind of a consideration prevents the test from being made up entirely of *Same* items. If all were *Same* items, there probably would be a point where persons who know the paired concepts would give their attention to over-all strategy instead of responding to the items. For example, persons knowing most of the items might be inclined to favor the *Different* response in order to keep a reasonable proportion of both kinds of answers, while others might guess the composition of the test and score them all as *Same*.

SUMMARY

Characteristics of response below the manifest level of chance on a verbal test (*VPT*) are examined and are indicated to be meaningful. Considerations of strategies of response to individual items and to the test as a whole are discussed.

REFERENCES

1. BORGATTA, E. F., and CORSINI, R. J. *Quick Word Test*. Yonkers-on Hudson, New York: World Book Company, 1960.
2. FRANCESCO, E. The Verbal Power Test of Concept Equivalents (*VPT*). *J. Psychol.*, 1960, *49*, 213-216.

SOCIAL DESIRABILITY AND THE TYPE OF HOSTILITY EXPRESSED ON THE TAT

MARVIN REZNIKOFF AND ADELAIDE DOLLIN

Institute of Living, Hartford, Conn.

In a recent investigation, Allison and Hunt [1] studied the relationship between aggression and the Edwards social desirability (SD) variable [3]. Positing that verbal aggression is generally seen as undesirable in our culture, they tested the proposition that high SD people as measured by the Edwards Scale would be likely to express less aggression than low SD people. Utilizing a specially devised paper-and-pencil "Situational Frustration Test", these investigators obtained a significant inverse association between amount of aggression elicited on this test and independent SD ratings. They concluded that their results could be taken to indicate construct validity of the Edwards SD concept.

While Allison and Hunt focussed entirely on quantity of aggression and SD, the present study examined amount and another factor of potentially equal importance —the *type* or *kind* of aggression in relation to SD. In this regard it could be conjectured, for example, that individuals with a strong SD orientation might be basically quite hostile but consistently manifest their hostility in a less direct manner. Specifically, then, expressions of overt and covert hostility were compared in a group of subjects categorized according to their scores on the Edwards SD Scale.

Procedure

Forty-eight psychiatric inpatients falling into a variety of non-organic psychodiagnostic categories participated in this study. All were of average intelligence or better and had at least a year of high school. An Edwards SD Scale and six TAT cards were administered to each. The six TAT pictures were selected to sample a range of emotional tone, two being rather happy or pleasant (2 and 10), two unhappy or sad (3 BM and 13 MF), and the remaining two (1 and 4) intermediate in their connotation[2].

Each of the six TAT stories collected from the forty-eight subjects was evaluated for hostile content by two judges working individually and utilizing a hostility scale developed by Hafner and Kaplan[4]. This latter procedure, which was modified slightly for the current study, enumerates and classifies TAT themes as indicative of either overt or covert hostility and assigns weights to them on a four point scale. A reliability coefficient of .90 was obtained between the independent ratings of the two judges, employing the Hafner and Kaplan criteria of aggression.

Results and Discussion

On the basis of their Edwards SD scores the subjects were divided into four equal groups. The ratings of the two judges were pooled and two TAT cards each combined under the three major categories of happy, neutral and sad. A double classification analysis of variance design with replicates was then used to compare the quantities of hostility found in four SD groups, as revealed in their responses to the three pairs of TAT cards.

As can be seen from Table 1 the only significant differences obtained were in comparisons of the TAT card categories. The four SD groups did not differ in amounts of overt, covert or total hostility expressed. Neither were there any significant interactions between the TAT stimulus values and the SD groups.

Table 1. Analyses of Variance of Hostility Ratings

Analysis	S of V	df	MS	F	P
Overt hostility	Groups	3	7.60	1.85	n.s.
	TAT	2	56.71	13.83	.001
	Groups × TAT	6	.99	.24	n.s.
	Error	132	4.10		
Covert hostility	Groups	3	1.27	.50	n.s.
	TAT	2	50.02	19.54	.001
	Groups × TAT	6	4.45	1.74	n.s.
	Error	132	2.56		
Total hostility	Groups	3	10.34	2.11	n.s.
	TAT	2	202.76	41.46	.001
	Groups × TAT	6	4.98	1.02	n.s.
	Error	132	4.89		

The second and more important analysis for purposes of the paper, involved evaluation of type of hostility manifested. Accordingly a sign test[5] was performed, ranking overt and covert hostility with respect to each other for each subject in the

Table 2. Sign Test Applied to Overt and Covert Hostility in Extreme SD Groups (O = Overt; C = Covert)

	Distribution of Subjects				
Group	O > C	O = C	C > O	Total	P
Low SD	4	3	5	12	.50
High SD	2	1	9	12	.03

two extreme SD groups. The data deriving from this procedure are reported in Table 2. It can be seen that in the low SD group five of twelve patients expressed more covert than overt hostility with three expressing equal proportions of each—a finding strictly explainable on the basis of chance. This, in opposition to the performance of the high SD group in which nine of twelve subjects exhibited more covert than overt hostility, which significantly departs from chance expectations.

Certain of the current study's results are not in agreement with those reported by Allison and Hunt, indicating that subjects receiving high scores on the Edwards SD Scale express significantly less aggression than subjects with low SD ratings. The fact that the earlier investigation used college undergraduates and student nurses while the present population was composed of psychiatric patients may partially account for the disparity.

Most likely, however, the findings are at variance with each other primarily because of the nature of the two tests employed to elicit aggression. The frustration test developed by Allison and Hunt was a minimally disguised, comparatively direct measure of aggression utilizing multiple choice responses. Such a procedure would doubtless tend to sharpen existing disparities between SD groups. The TAT, by contrast, a projective technique, involving relatively ambiguous stimulus material not explicitly designed to evoke aggressive feelings might, in effect, be undercutting the operation of an SD set governing hostile expression. On the other hand, the finding that high SD subjects are significantly more covert than open in their aggression while low SD individuals are not, would seem to be in line with general SD theory. This aspect of the data would also emphasize the importance of recognizing type as well as quantity of aggression as a variable associated with SD orientation.

Summary

The Edwards Social Desirability (SD) Scale and a six card TAT were administered individually to each of forty-eight psychiatric inpatients with the principal object of determining whether the kind of aggression expressed, covert or overt, was related to SD ratings. Based on their SD scores, the subjects were separated into four equal groups and their TAT evaluated by two judges for quantity and type of hostility utilizing a procedure developed by Hafner and Kaplan specifically for this purpose. Analyses of variance revealed no significant differences among the four groups with regard to amounts of covert, overt and total hostility. A sign test, however, comparing overt with covert aggression for each subject in the two extreme groups, indicated that a significantly greater number of individuals than chance in the high SD group rated as having more covert hostility, while in the low SD group no consistent direction was evident. The findings were contrasted with an earlier study of aggression and SD, and briefly discussed.

References

1. Allison, J. and Hunt, D. Social desirability and expression of aggression under varying conditions of frustration. *J. consult. Psychol.*, 1959, *23*, 528-532.
2. Dollin, Adelaide. The effect of order of presentation on perception of TAT pictures. Unpublished doctoral dissertation. Univ. Connecticut, 1960.
3. Edwards, A. L. *The social desirability in personality assessment and research*. New York: Dryden Press, 1957.
4. Hafner, A. J. and Kaplan, A. M. Hostility content analysis of the Rorschach and TAT. *J. proj. Tech.*, 1960, *24*, 137-143.
5. Siegel, S. *Nonparametric statistics for behavior sciences*. New York: McGraw Hill, 1956.

A LONGITUDINAL COMPARISON OF INTELLIGENCE INVOLVING THE WECHSLER-BELLEVUE I AND WAIS

DONN R. DUNCAN AND ALBERT M. BARRETT

University of Florida

Problem and Procedure

In an intelligence testing course in 1950, the WB I was given under supervision to 28 unselected subjects ranging in age from 18 to 72 and varying widely in educational and occupational backgrounds. These same subjects were located in different parts of Florida, North Carolina, and Alabama and retested with the WAIS in 1960. In effect, this ten year interval made possible a longitudinal study. Admittedly the N is small and the group is varied, as no replication design had been planned. However, since longitudinal studies involving the same subjects tested with WB I and WAIS are rare, it is interesting to note the results for whatever contribution they might make to the problems of IQ constancy, decline of intelligence with age, and comparisons of two well known tests.

Results

In Table 1 the 1950 group tested with WB I, ranging in age from 18 to 72 years with an average age of 30.57, is compared with the same group tested with WAIS ten years later in 1960 when the age range was 28 to 82 and the average age was 40.57. The group was composed of 16 males and 12 females; no significant sex difference between mean IQ's was found. The range of IQ's on WB I for 1950 was 105 to 133 and for WAIS in 1960, 97 to 137.

Table 1 shows that the group means for both tests are high, indicating that the group is mostly composed of bright to superior people. Even though this is a rather selective group, the IQ's remained reasonably constant over the ten year period.

Table 1. Comparison of WB-I and WAIS, Full Scale, Verbal Scale, Performance Scale and Subtest Scaled Scores (N = 28).

Subtests	WB I 1950		WAIS 1960				WB I/WAIS
	Mean	SD	Mean	SD	*t*	P	*r*
Full Scale	118.57	8.24	119.29	8.81	.72		.82
Verbal Scale	117.71	7.98	120.86	9.32	3.14	.01	.83
Information	12.57	1.84	13.79	2.06	3.76	.01	.64
Comprehension	12.25	2.26	14.50	2.57	4.22	.01	.35
Arithmetic	12.71	3.44	12.64	2.83	0.13	...	.58
Similarities	12.93	2.17	13.89	2.11	2.16	.05	.41
Digit Span	10.46	2.67	11.11	3.53	1.24	...	.69
Vocabulary	12.54	1.78	13.81	2.34	3.89	.01	.72
Performance Scale	116.18	9.88	115.18	8.54	.54		.45
Digit Symbol	11.93	2.67	11.00	3.00	2.72	.02	.81
Pict. Comp.	11.50	2.11	11.96	2.18	0.69	...	.38
Block Design	13.81	2.54	12.25	3.21	1.14	...	.62
Pict. Arr.	10.96	2.73	11.00	2.80	0.05	...	.18
Obj. Assem.	11.36	3.21	10.50	2.43	1.72	.10	.49

Although the results obtained are in some degree comparable to a few previous studies of WB I and WAIS such as those of Cole and Weleba[1], Goolishian and Ramsay[3], and Dana[2], they are not directly parallel because the earlier studies were not longitudinal in character.

Subtest score comparisons were made and are presented in Table 1. Digit Symbol, Vocabulary and Digit Span have the highest correlations. Both Vocabulary and Digit Symbol also showed significant differences between means. The signif-

cant changes between means of some of the subtest scores are probably attributable to differences in standardizing procedures between WB I and WAIS.

Conclusions

A group of 28 unselected subjects was tested in 1950 with the Wechsler Bellevue I and retested in 1960 with the WAIS. The IQ's remained relatively constant for the group over a ten year period. This finding would appear to attest to the validity of the age correlation factors and the readjustment of WAIS scores formulated by Wechsler.

References

1. Cole, D. and Weleba, Lois. Comparison data on the Wechsler-Bellevue and the WAIS. *J. clin. Psychol.*, 1956, *12*, 198-199.
2. Dana, R. H. A comparison of four verbal subtests on the Wechsler-Bellevue, form I, and the WAIS. *J. clin. Psychol.*, 1957, *13*, 70-71.
3. Goolishian, H. A. and Ramsay, Rose. The Wechsler-Bellevue, form I, and the WAIS: a comparison. *J. clin. Psychol.*, 1956, *12*, 147-151.

PSYCHOLOGICAL EFFECTS OF ISOCARBOXAZID AND NIALAMIDE ON A GROUP OF DEPRESSED PATIENTS

VASSO VASSILIOU AND H. E. HIMWICH

Thudichum Psychiatric Research Laboratory
Galesburg (Ill.) State Research Hospital

Problem

The present investigation compared the effects of isocarboxazid* (Marplan, Hoffman-LaRoche) and nialamide* (Niamid, Pfizer), both monoamine oxidase inhibitors, on a group of hospitalized depressed patients in their performance on a selected battery of tests. Although the effects of these drugs are not known, it was suggested that the drugs have the property of reducing depressive symptoms. However, it was thought interesting to determine the effects of the drugs on other areas of functioning on an exploratory basis. Tests involving a variety of areas, such as ideational processes, psychomotor coordination, and mood were included in the test battery.

Method

Tests. The main author had no knowledge of the suggested effects of the drugs and consequently a variety of tests were selected to determine what the effects of these drugs were. Eleven tests were chosen to assess changes in ideational processes, psychomotor coordination and mood. Four tests involving the Vocabulary, Digit Span, five problems from the Arithmetic subtest, and Digit Symbol were chosen from the Wechsler Adult Intelligence Scale. Other tests included: a Visual Design Memory test; an Immediate, Intermediate and Past Memory test; a Visual Object Memory test; and selected items from Sacks Sentence Completion test.

Scoring. All tests, except the Sentence Completion test, were of an objective nature and did not present quantitative scoring problems. The Sentence Completion data were rated on eight bipolar scales with seven intervals in an attempt to differentiate "Drive Orientation" and "Personal Assessment" under the three experimental conditions. The scales under "Drive Orientation" were: *strong-weak, positive-negative, appropriate-inappropriate;* those under "Personal Assessment" were: *positive-negative, happy-sad, active-passive, satisfied-unsatisfied, acceptable-unacceptable.* The same scales were used for each of three testings. Three psychologists rated the Sentence Completion test independently on the eight bipolar scales. To determine the degree of agreement among the three sets of independent scores the Kendall coefficient of concordance was used. Reliability coefficients secured for "Drive Orientation" were: .82, .83, .88 respectively and for "Personal Assessment" .80, .87, .85. Any disagreement was resolved through conference.

Subjects and Procedure. Thirty hospitalized patients involving 16 females and 14 males selected on the judgment of three psychiatrists on the basis of depressive symptoms were the subjects (Ss). The mean age of the females was 48.64, (range 20 to 68) and the mean age for the males was 50 (range 34 to 72). Initially all Ss were able to participate in the testing and work on the

*We are grateful to Hoffman-LaRoche, Inc., Nutley, N. J., and to Pfizer Laboratories of Groton, Connecticut for the generous supplies of isocarboxazid and nialamide respectively.

tasks assigned to them with the exception of two males and two females who refused to take the tests. Thus four Ss were eliminated from this study. The Ss were originally tested under condition one (C1), after being placed on Placebo for eight weeks without administration of the drug. The second testing involved condition 2 (C2) following the administration of the drug eight weeks later. The male patients were on isocarboxazid (Marplan) starting on 10 mg T.I.D. and the dosage was gradually increased until the patients responded, but not higher than a maximum of 20 mg T.I.D. The female patients were on nialamide (Niamid). They started on 25 mg T.I.D. and went to a maximum of 75 mg T.I.D. Both drugs are monoamine oxidase inhibitors. Differences in dosage were due to the fact that isocarboxazid is more potent than nialamide. Dosages were arranged in such a way in order to secure equal clinical potency. Following that the Ss were seen and tested for a third time after a period of eight weeks during which all Ss were placed on placebo again and this was called condition 3 (C3). Neither the Ss nor the main investigator were aware of the actual medication until the completion of the study.

The 14 female Ss and 12 male Ss were combined and treated as a single sample in the analysis of the data because no significant differentiations between the two groups in their scores on the test material were observed. Comparison between the male and female Ss by the Mann Whitney nonparametric technique produced no significant differences.

Results

With the exception of Sentence Completion test, each test from the battery was analyzed individually under the three conditions specified in the procedure. Because of the apparent skewness of the scores, the results were analyzed by the nonparametric Friedman two-way analysis of variance by ranks[1] to determine whether the rank totals differ significantly with results presented in Table 1.

Table 1 reveals that all subtests were significant beyond the .01 level. In every case the score for C1 was the largest, followed by the score in C3. C2, when drug was introduced, revealed the lowest score. Lower scores in all subtests indicate a poorer performance.

Table 1. Mean Scores and the Friedman Two-Way Analysis of Variance on the Ten Different Subtests under the Three Experimental Conditions

Test	C1	C2	C3	χ^2
Vocabulary	29.5	20.34	23.69	29.3*
Immediate Memory	8.19	5.38	5.96	29.2*
Intermediate Memory	24.23	18.84	20.00	28.8*
Past Memory	18.84	11.92	14.61	29.0*
Visual Object Memory	59.61	50.38	56.92	26.8*
Visual Design	36.92	26.92	31.15	28.0*
Digit Span Forward	48.46	37.69	41.53	27.1*
Digit Span Backwards	34.23	25.00	25.38	27.5*
Arithmetic	27.30	21.53	24.61	27.0*
Digit Symbol	212.30	146.15	198.46	28.4*

*Significant beyond the .01 level

From the Sentence Completion test, although reliabilities were high, results were not analyzed for discussion because inspection of the test data do not reveal any changes in the areas of "Drive Orientation" and "Personal Assessment".

Summary

The purpose of this study was to explore the effects of isocarboxazid (Marplan) and nialamide (Niamid), both monoamine oxidase inhibitors, on depressed patients. Eleven tests were chosen to assess changes in ideational processes, psychomotor coordination and mood. Patients were tested under three experimental conditions. It was found that isocarboxazid and nialamide lowered the performance of the Ss in the tests involving functions referring to higher mental processes and psychomotor coordination.

References

1. Siegel, S. *Nonparametric Statistics.* McGraw-Hill Book, New York, 1956.

A CROSS-VALIDATION OF LEARY'S "PUBLIC" COMMUNICATION LEVEL

WALTER G. KLOPFER

University of Portland

Problem

Leary[1] has specified operational measures of five levels of interpersonal behavior. These include the Level of Public Communication (the way in which an individual is perceived by people in general), the Level of Conscious Perception (the way in which he perceives himself), the Level of Private Symbolization (the way he is evaluated on the basis of projective tests), the Level of the Unexpressed Unconscious (significant omissions at the preceding levels), and the Level of Values and Goals (perception of the ideal person). Of all these, the first, or "Public" level, has been the most generally ignored in typical personality assessment programs, due largely to the difficulty of attacking it directly in a cross-sectional evaluation of an individual. Nevertheless, it would seem to be a vital part of any comprehensive method of evaluating interpersonal behavior since it will lead directly to predictions of social stimulus value, which in turn will enable the diagnostician to evaluate the tendency on the part of the individual being assessed to provoke various kinds of reactions on the part of other people. Thus, a person who at the "Public" level demonstrates dominance is likely to provoke submission; if he demonstrates aggression he is likely to provoke rebellion; if he demonstrates docility he is likely to provoke responsible, managing behavior, etc.

Leary's method of predicting such "public" behavior is to use formulae derived from various scales of the MMPI. The two main dimensions of his interpersonal system are Dominance-Submission and Love-Hate, and the two together form coordinates on which a summary point can be plotted. These summary points lead to more specific predictions of habitual interpersonal behavior as it is perceived by people in general. The question being attacked in this study is whether the predictions based upon the MMPI can be cross-validated by using an independent predictor of "Public" interpersonal behavior along the two dimensions specified in Leary's system.

Procedure

Thirty-one outpatients seen in a university clinic were given a complete battery of psychological tests, including the MMPI, and an interview. After evaluating all the material, with the exception of the MMPI, the psychologist was asked to fill out a copy of Leary's Interpersonal Adjective Check List on the patient in terms of how he is likely to be perceived by people in general. Thus, the clinician is asked to give a prediction of Public Level behavior of the patient, based upon a multi-level assessment procedure. The Adjective Check List was scored according to Leary's standard system, and measures of Dominance and Love on each individual thus became available. The MMPI scales were then analyzed and predictions made according to Leary's usual system on the same two dimensions. The group of patients was then paired in terms of the Love and Dominance dimensions as to how consistently their "Public" behavior was being predicted on the MMPI and the IACL.

Results

The Pearson product moment correlation of Dominance as measured by the MMPI and IACL was .29, which is not significantly different from chance. A similar correlation of Love, however, yielded a correlation of .54, significant at the .01 level. It would be of major value if a standard psychological instrument like the MMPI which is self-administering and scorable clerically, could indeed produce valid predictions of "Public" behavior.

SUMMARY

This study has tended to validate Leary's system of predicting "Public" behavior from MMPI formulae in part, in that Affection and Hostility are apparently predictable validly from the MMPI as judged by the criterion herein employed. However, the Dominance dimension fared less well. It is difficult to know from the results of the present study whether this is because the Dominance dimension lends itself less easily to prediction by clinicians on the basis of a standard clinical experience, or whether indeed Leary's system is weak in part. Further research is indicated to refine further the degree to which the MMPI can be used to predict "Public" behavior in the manner suggested by Leary. This might take the form of deriving and cross-validating new empirical scales of Love and Dominance from the MMPI on the basis of *S*s judged clinically on these dimensions.

REFERENCES

1. LEARY, T. F. *The Interpersonal Diagnosis of Personality.* New York: The Ronald Press Co., 1957.

CHANGES IN ESSENTIAL HYPERTENSION OVER TIME: A FACTOR ANALYTIC CASE STUDY*

G. A. MILTON

University of Colorado

Early in the literature on hypertension, Alexander demonstrated a relation between the "emotional state" of an hypertensive patient and changes in his blood pressure over time.(1, p. 311) Specific emotional states, behaviors, and personality characteristics have also been attributed to the hypertensive patient as differentiated from patients with other psychosomatic disorders and non-patient populations.(1, p. 292; 2, p. 746; 3) The most often noted characteristics are repressed hostility and conflict about sex-role identification. To the extent that these and other emotional states are important in the elevations in blood pressure of an hypertensive it is to be expected that correlations would be obtained between fluctuations in blood pressure and the specific variables. The present study investigates the relationships over time among blood pressure, personality dynamics discussed during therapy, and overt behaviors for a single hypertensive patient.

THE PATIENT

The patient was a 38 year old white male hypertensive who was still in the fluctuating stage of hypertension. His blood pressures ranged between 120/70 and 140/200 during the nine months period in which he was followed. He was hospitalized in a G M and S hospital during this period and was receiving biweekly psychotherapy. Although he had a mixed psychiatric diagnosis with both characterological and obsessive-compulsive features, he displayed a classic hypertensive syndrome from Dunbar(2, p. 746) including problems with hostility, authority, especially where women were concerned, lack of full utilization of his intelligence, problems of impotence, and an interest in playing golf.

*The author is indebted to Dr. Jack Blumenkrantz and Mrs. Dorothy Frey Marshall for their assistance on this project.

Procedures

Four groups of variables are used in this analysis: one measure of blood pressure, nine measures based on content analyses of psychotherapy interviews, four measures based on records of the patient's behavior on the ward, and one measure of the amount of drugs being taken by the patient. The inter-correlations among these 15 variables as measured over the nine-month period were then factor analyzed by the centroid method, and the resulting centroid factors were rotated orthogonally until what appeared to approximate simple structure was achieved. The patterns which emerged are discussed in the results section.

The patient's blood pressure was measured daily by the medical staff over the period of this study, but the psychotherapist avoided any knowledge of this variable in order to prevent any methodological contamination between this variable and the variables measured in the therapy sessions. Mean blood pressure

$$\frac{(2 \text{ Diastolic} + \text{Systolic})}{3}$$

was used as the score in this analysis.

The patient was seen in 75 biweekly therapy sessions over the nine-month period. The therapist did not introduce topics of discussion. During and after each therapy hour the therapist made detailed descriptive notes of the session which were subsequently rated by two judges for the presence of the following variables. (The patient had vigorously objected to the use of a tape recorder.) The inter-rater reliabilities for each variable are presented after the title of the variable.

2. *Hostility* (.88). Any discussion of the patient's feelings of anger or hostile acts.
3. *Identification* (.70). Discussion of the patient's conflict about role identity, activity-passivity, what he wants to be.
4. *Sex* (.78). Discussion of sexual behavior and impulses.
5. *Manipulation* (.81). Discussion of the patient's attempts at controlling others or his environment.
6. *Physical comfort* (.88). Discussion of health, physical well-being, comfort.
7. *Authority* (.67). Discussion of authority problems, rules, superior-peer relations.
8. *Family Origins* (.93). Discussion of childhood, home and family.
9. *Dreams* (.99). The patient relates a dream and works with it.
10. *Patient-Therapist Relations* (.77). Discussion of the therapy interaction and feelings about it.

Notes about the patient's behavior on the ward by the nursing staff as they appeared in the chart were recorded for the following variables:

11. *Tension*, when the patient was noted as being tense and upset.
12. *Hostile behavior*, when the patient displayed such behavior.
13. *Complaints*, when the patient complained about service, conditions, etc.
14. *Acting out*, when the patient broke rules, tested limits, and failed to keep appointments.

The last variable included in the analysis is the amount of drugs being given the patient by the ward physician. This is included primarily as a control variable since the dosage ranged rather widely over the nine-month period.

Results

The inter-correlations among the 15 variables are presented in Table 1. Inspection of the matrix suggests several clusters of variables within this patient. For example, we find that when the patient's blood pressure was high he was more inclined to discuss dreams and identification problems, he was more tense and less hostile on the ward, and he was receiving smaller drug dosages. These patterns become clear when the results of the factor analysis are examined. Five factors emerged

TABLE 1. INTERCORRELATIONS AMONG THE 15 VARIABLES

Variable 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	02	54	−09	35	24	−02	28	70	39	21	−29	−04	10	−25
2		10	−23	47	13	64	03	21	50	−21	09	−43	−10	−37
3			−04	03	07	39	71	42	32	03	−34	−22	−06	−23
4				−12	51	−14	04	43	−36	16	−57	−08	20	28
5					10	10	−32	30	63	−11	20	15	15	−24
6						07	35	28	−28	23	−04	52	38	03
7							40	−16	14	00	07	−48	−23	−01
8								74	−25	00	01	04	−25	07
9									06	13	04	06	01	−24
10										−13	−03	13	11	−18
11											09	27	49	31
12												06	09	19
13													75	32
14														16
15														

from the analysis and the rotated factor matrix is presented in Table 2. All loadings of. 35 and above have been marked with an asterisk to indicate that they probably have a stable relationship to the factor. All 5 factors seem to be meaningfully defined.

TABLE 2. ROTATED FACTOR MATRIX: HYPERTENSIVE PATIENT FACTORS

Variable	A	B	C	D	E
1. Mean Blood Pressure	.40*	.25	.81*	−.09	.21
2. Hostility	−.22	.63*	.04	.18	.22
3. Identification	−.11	.26	.78*	.17	.21
4. Sex	.28	−.50*	.08	.43*	.28
5. Manipulation	.21	.70*	−.05	−.19	.13
6. Physical Comfort	−.13	.42*	.07	.16	.70*
7. Authority	.00	.01	.00	.81*	−.12
8. Family Origins	.02	.07	.31	.28	.56*
9. Dreams	.55*	.12	.24	−.10	.82*
10. Patient-Therapist Relations	.06	.71*	.12	−.13	−.12
11. Tension	.36*	−.01	.07	.21	−.10
12. Hostile Behavior	.41*	.28	−.41*	.07	.08
13. Complaints	.76*	.02	−.19	.04	−.09
14. Acting Out	.75*	.21	−.11	.19	−.42*
15. Drugs	−.19	−.19	−.48*	.05	−.20

*Significant loading.

INTERPRETATION OF THE FACTORS

Factor A seems to be a classic hypertensive pattern as it has emerged in this individual. This seems to be an *emotional state* factor.

Variable	Loading
13. Complaints	.76
14. Acting Out	.75
9. Dreams	.55
12. Hostile Behavior	.41
1. Blood Pressure	.40
11. Tension	.36

This pattern is of considerable significance for psychosomatic theory. In this patient we find complaining, rule breaking, hostile, tense behavior varying over time with fluctuations in blood pressure. However, we do not find that his discussion of hostility is associated with this pattern. In fact, during these periods of time, he seems less inclined to discuss hostility (−.22) which would be quite consistent with the theoretical position that the *repression* of hostile ideation is essential to the elevation of blood pressure in a hypertensive.

Factor B may best be described as an *interpersonal relations factor.*

	Variable	Loading
10.	Patient-Therapist Relations	.71
5.	Manipulation	.70
2.	Hostility	.63
6.	Physical Comfort	.42
4.	Sex	−.50

All of the variables loaded on this factor are from the therapy material and suggest a consistent pattern of associations in the patient's thought. In discussing interpersonal relations he tends also to be concerned with his own well-being and hostility, but does *not* associate this with sexuality. Sex, for this patient, has a separate and distinct meaning as will be seen in factor D.

Factor C points out quite clearly the significance of the *identification conflict* in the blood pressure fluctuations of this patient.

	Variable	Loading
1.	Blood Pressure	.81
3.	Identification	.78
12.	Hostile behavior	−.41
15.	Drugs	−.48

This factor is also of importance for psychosomatic theory. The fluctuations in this patient's blood pressure seem associated to a marked degree with his concern about identification. It will also be noted that during periods of elevated blood pressure and concern about identification he is not showing manifest hostility. That the amount of drugs being taken by the patient is also negatively loaded in this pattern suggests that the drugs did, in fact, help control his blood pressure, but that they also had the effect of reducing his motivation to talk about this central conflict area of identification.

Factor D seems simply to illustrate a *sex-authority* linkage in this patient.

Variable	Loading
Authority	.81
Sex	.43

Authority problems are associated with sexuality and the sexual act for this patient. The content of these discussions frequently involved frustration and impotence associated with both spheres.

Factor E is a pattern of less clarity than the others but may indicate a *reflective introspective attitude* on the part of the patient.

	Variable	Loading
9.	Dreams	.82
6.	Physical comfort	.70
8.	Family origins	.56
14.	Acting out	−.42

That this pattern emerges suggests that the patient had periods of time when he was actively working on the problems of his well-being, seeking their origin in his early experience. During these periods he appears to have been "well behaved" in the sense of conforming to the rules of his hospital society.

Discussion

Of the five factors which emerged in this analysis at least two are of considerable significance for psychosomatic theory. For this patient, fluctuation in blood pressure over time is clearly related to patterns involving hostility and identification. This is supportive of the inference drawn from previous research which indicates that hypertensive patients, as a group, have greater problems in these areas than do

other patient groups or non-patient groups[1, 2, 3]. However, for this patient, the emerging pattern was not one combining hostility and identification, but rather two separate patterns. One (Factor A) is a pattern of emotional disturbance and active overt acting out of the hostility without any evidence of awareness of the significance of this behavior. The other is a consideration of problems of identity, accompanied by, in fact, lowered manifest hostility. This clearly supports Alexander's assertion that in the hypertensive the fluctuations in blood pressure are a direct result of fluctuations in the hostile impulses[1, p. 292]. Also this indicates that the identification problems which Kalis, Harris, et al.[3] emphasize in connection with the hypertensive and which Alexander sees as a "characteristic psychodynamic structure" [1 p. 292], have a direct and fluctuating covariance with elevations in the patient's blood pressure.

Of the three other patterns which emerged at least one, the sex-authority factor (D), is also referred to in the literature[2, p. 746]. This factor, however, is not related to the fluctuation in blood pressure and, for this patient at least, must be regarded as an independent aspect of his personality. There is also evidence that this patient linked hostility and manipulation with interpersonal relations (Factor B) and that he had definite periods of seeking an answer in his early experiences (Factor E) but these patterns, also, did not covary with his blood pressure.

Summary

This study is a factor analysis of 15 variables measured over time for a single hypertensive patient. The hypothesis investigated is that emotional and behavioral variables which are significant in fluctuations of the patient's blood pressure will emerge from this analysis. Five factors were isolated, two of which were loaded with the blood pressure measure. These two seem significant for psychosomatic theory since one of the factors is related to overt hostility and the other to problems of identification.

References

1. Alexander, F. and French, T. M. (eds.) *Studies in psychosomatic medicine.* New York: Ronald Press, 1948.
2. Dunbar, H. F. *Emotions and bodily changes.* New York: Columbia University Press, 1954.
3. Kalis, B. L., Harris, R. E., Bennett, L. F., and Sololow, M. "Personality and the Predisposition to essential hypertension." Paper presented at the annual meeting of the American Psychological Association, August, 1956.

EDITORIAL OPINION

THE FUTURE OF APA DIVISION 12

The Winter 1961 *Newsletter of the Division of Clinical Psychology* contains a letter by Victor Raimy suggesting a reorganization of Division 12 to include within it an "American Society of Clinical Psychology" with membership open to Ph.D.'s professionally qualified to practice clinical psychology. The *raison d'etre* of the new ASCP would be to provide the mechanism for a strong social action group in our rapidly coming-of-age profession which is increasing concerned with its own impetus and enormous potential as a profession.

There is some question whether simply forming an ASCP within Division 12 consisting of the same persons who now dominate Division 12 membership would solve the issues which have resulted in dissatisfaction with Div. 12 in its present composition. There is a great danger of creating an *elite guard* within clinical psycho-

logy comparable to the GROUP FOR THE ADVANCEMENT OF PSYCHIATRY (GAP) which created such dissension in the AMERICAN PSYCHIATRIC ASSOCIATION a few years ago when it constituted itself as a self-appointed action group to guide psychiatry. We see no purpose in duplicating an elite group whose membership and functions would be overlapping with Div. 12 in important respects.

The real split within Div. 12 exists between the academic professors who have dominated the Division for years and the practitioners in the field who want a truly professional association to represent them. This phenomenon is not limited to clinical psychology but has existed in other clinical professions where the needs of basic science professors and practicing clinicians have differed greatly. In clinical psychology there has never been a strong association of professional clinicians who derive a major portion of their income from actual practice. For many reasons, there has been a tendency of academicians to look down upon full time clinicians and upon those in private practice. Conversely, there has been some tendency among practitioners to resent academicians. In the history of the Division of Abnormal and Clinical Psychology, there has been a tendency for the academicians to be interested primarily in abnormal psychology, and for the practitioners to be more interested in the professional promotional activities of the division. There is serious question whether the two interest groups belong together and whether their joint interests would be better served by having two separate professional organizations.

In our opinion, Division 12 might better be left in its present form as a special interest group within the AMERICAN PSYCHOLOGICAL ASSOCIATION serving the same functions as the specialty sections of the AMERICAN MEDICAL ASSOCIATION. We believe that the proposed AMERICAN SOCIETY OF CLINICAL PSYCHOLOGY should be organized independently of the APA and to represent primarily the interest of practicing clinicians as their social action group. The ASCP would thus function in the same independent capacity as does the AMERICAN PSYCHIATRIC ASSOCIATION having its own organization and functions.

Personally, we are unwilling to see any substantial increase in the dues of Div. 12 simply to create another bureaucracy which does not solve the basic issues at stake. For Div. 12 to set up its own executive secretary and offices would be the first step in duplicating functions which properly belong to the APA and which should not come into competition with the APA. There is some question as to what faction in the present Div. 12 such an executive secretary and office would represent. Certainly, there is no reason to believe that it would fairly represent the practicing clinicians who are clamoring for representation. And here, we would like to make a suggestion whereby the practicing clinicians could be assured better representation in Div. 12 business. It is simply the suggestion that academicians and practitioners should alternate in the Presidency and important appointments within the Division so that all groups might have representation.

We would like to secure opinions from Division 12 members concerning the feasibility of these suggestions and would be glad to have an informal discussion of these alternatives at the coming 1961 Business Meeting of Division 12.

F. C. T.

NEW 1961 BOOKS IN PSYCHOLOGY

☐ *Dominick A. Barbara*—**THE PSYCHOTHERAPY OF STUTTERING.** Pub. Sept. '61, 304 pp., 18 il.

☐ *Jack Block*—**THE Q-SORT METHOD IN PERSONALITY ASSESSMENT AND PSYCHIATRIC RESEARCH.** Pub. Aug. '61, 144 pp. (Amer. Lec. Psychology)

☐ *Ted Brown*—**THE ENIGMA OF DRUG ADDICTION.** Pub. Aug. '61, 352 pp., 49 il. (Police Science Series)

☐ *William J. Bryan, Jr.*—**RELIGIOUS ASPECTS OF HYPNOSIS.** Pub. Oct. '61, 112 pp., 1 il.

☐ *H. S. Burr*—**THE NEURAL BASIS OF HUMAN BEHAVIOR.** Pub. Dec. '60, 272 pp., 8 il., $7.50

☐ *Arthur C. Carr et al.*—**THE PREDICTION OF OVERT BEHAVIOR THROUGH THE USE OF PROJECTIVE TECHNIQUES.** Pub. Dec. '60, 192 pp., 23 il. (Amer. Lec. Psychology), $6.75

☐ *C. J. Ducasse*—**A CRITICAL EXAMINATION OF THE BELIEF IN A LIFE AFTER DEATH.** Pub. Feb. '61, 336 pp. (Amer. Lec. Philosophy), $875.

☐ *Malinda Dean Garton*—**TEACHING THE EDUCABLE MENTALLY RETARDED—PRACTICAL METHODS.** Pub. March '61, 248 pp., 147 il., $7.50

☐ *Andrew Gemant*—**THE NATURE OF THE GENIUS.** Pub. May '61, 216 pp., $6.50

☐ *R. W. Gerard et al.*—**BRAIN MECHANISMS AND LEARNING.** Pub. Aug. '61, 648 pp., 245 il.

☐ *J. C. Glidewell*—**PARENTAL ATTITUDES AND CHILD BEHAVIOR.** Pub. June '61, 272 pp., 18 il., $8.50

☐ *Milton Greenblatt*—**MENTAL PATIENTS IN TRANSITION.** Pub. July '61, 408 pp., 16 il.

☐ *Molly Harrower*—**THE PRACTICE OF CLINICAL PSYCHOLOGY.** Pub. Oct. '61, (Amer. Lec. Psychology)

☐ *Louis Hausman*—**ILLUSTRATIONS OF THE NERVOUS SYSTEM. Atlas III.** Pub. April '61, 208 pp. (8 3/8 x 10 7/8), 546 il., $9.50

☐ *Irving Louis Horowitz*—**PHILOSOPHY, SCIENCE AND THE SOCIOLOGY OF KNOWLEDGE.** Pub. Jan. '61, 192 pp. (Amer. Lec. Philosophy), $7.75

☐ *Jerome Kagan and Gerald S. Lesser*—**CONTEMPORARY ISSUES IN THEMATIC APPERCEPTIVE METHODS.** Pub. March '61, 352 pp., $12.00

☐ *Alfred H. Katz*—**PARENTS OF THE HANDICAPPED: Self-Organized Parents' and Relatives' Groups for Treatment of Ill and Handicapped Children.** Pub. Jan. '61, 168 pp., $6.00

☐ *Nathaniel M. Levin*—**VOICE AND SPEECH DISORDERS.** Pub. Aug. '61, 944 pp., 450 il.

☐ *Arnold M. Ludwig*—**THE MYSTERY OF LIFE.** Pub. Sept. '61, 160 pp.

☐ *John M. Macdonald*—**THE MURDERER AND HIS VICTIM.** Pub. May '61, 436 pp., $10.50.

☐ *John M. Martin*—**JUVENILE VANDALISM:** A Study of Its Nature and Prevention. Pub. Feb. '61, 160 pp., 12 il., $6.50

☐ *M. F. Ashley Montagu*—**GENETIC MECHANISMS IN HUMAN DISEASE.** Pub. Oct. '61, 512 pp., 365 il.

☐ *M. F. Ashley Montagu*—**PRENATAL INFLUENCES.** Pub. Oct. '61.

☐ *Charles E. Palmer*—**SPEECH AND HEARING PROBLEMS.** Pub. May '61, 152 pp., 15 il., $5.50

☐ *J. H. Quastel and David M. J. Quastel*—**THE CHEMISTRY OF BRAIN METABOLISM IN HEALTH AND DISEASE.** Pub. March '61, 184 pp., 23 il. (Amer. Lec. Living Chemistry), $6.50

☐ *Robert N. Rapoport*—**COMMUNITY AS DOCTOR:** New Perspectives on a Therapeutic Community. Pub. April '61, 336 pp., 14 il., $9.75

☐ *Elizabeth Lodge Rees*—**A DOCTOR LOOKS AT TOYS.** Pub. Aug. '61, 208 pp., 31 il.

☐ *Albert Scheflen*—**A PSYCHOTHERAPY OF SCHIZOPHRENIA.** Pub. March '61, 304 pp., 1 il., $8.50

☐ *Alexander Simon et al.*—**THE PHYSIOLOGY OF EMOTIONS.** Pub. April '61, 260 pp., 20 il., $8.50

☐ *R. H. Smythe*—**ANIMAL VISION.** Pub. May '61

☐ *R. H. Smythe*—**ANIMAL PSYCHOLOGY.** Pub. Sept. '61

☐ *R. H. Smythe*—**HOW ANIMALS TALK.** Pub. Nov. '61, 110 pp., 10 il.

☐ *R. H. Smythe*—**THE MIND OF THE DOG.** Pub. Nov. '61, 118 pp., 7 il.

CHARLES C THOMAS • PUBLISHER 301-327 East Lawrence Avenue **SPRINGFIELD • ILLINOIS**

NEW TEST!!

"The new in tests . . . . . "

How would you like to use a projective personality test which can be scored quickly and objectively? Whose interpretation is based on careful factor-analytic studies? Which has norms for 11-year-old through adult levels (over 1900 selected *Ss*)? Uses an interesting and disarming cartoon-like format?

Cassel and Kahn

GROUP PERSONALITY PROJECTIVE TEST

Order Examiner's Set including general manual, scoring keys, 10 test booklets, and 100 answer sheets, enough material for a good-sized testing program (and only $13.50).

•

For further information, write to

PSYCHOLOGICAL TEST SPECIALISTS
Box 1441
Missoula, Montana

"The new in tests . . . . . "

NEW TEST!!

When the Mountain Did Go to Mahomet

For the parents of retarded or emotionally disturbed children, transportation expenses for enrollment and visits at a usually distant treatment center — added to the cost of residential treatment itself — could, on occasion, make it impossible for the parents to give their child the benefits of an individualized, twenty-four-hours-a-day program, under full professional guidance.

Cognizant of this factor, The Devereux Foundation has pioneered three branches, which, in effect, make it one of the most accessible residential treatment centers in the United States. At each branch outstanding therapeutic, educational, and vocational services are available.

Serving the East Coast, Devereux Schools are located at Devon, Pennsylvania (Mr. Charles J. Fowler, Director of Admissions); serving the West Coast, Devereux Schools at Santa Barbara, California (Mr. Keith A. Seaton, Registrar); and serving the Southwest, Devereux Schools at Victoria, Texas (Mr. John M. Barclay, Director of Development). These will welcome your inquiry.

THE DEVEREUX FOUNDATION

A nonprofit organization
Founded 1912
Devon, Pennsylvania
Santa Barbara, California
Victoria, Texas

SCHOOLS
COMMUNITIES
CAMPS
TRAINING
RESEARCH

HELENA T. DEVEREUX
Administrative Consultant

EDWARD L. FRENCH, Ph.D.
Director